Mathematical Aspects
of
Computer Engineering

Advances
in
Science
and
Technology
in
the USSR

Mathematics and
Mechanics Series

Mathematical Aspects of Computer Engineering

Edited by
V. P. MASLOV, Mem. USSR Acad. Sci.
K. A. VOLOSOV, Cand. Sci. (Phys.-Math.)

MIR Publishers
Moscow

Математические аспекты вычислительной техники
Под редакцией акад. *В. П. Маслова*
и канд. физ-мат. наук *К. А. Волосова*
Издательство «Мир» Москва

Translated from Russian
by Eugene Yankovsky

First published 1988

На английском языке

Printed in the Union of Soviet Socialist Republics

ISBN 5-03-000219-7

CONTENTS

PREFACE

The present collection of articles is the result of many years of research conducted by our team into various aspects of designing and building the component base of promising high-speed computational systems. The articles deal with the following topics:

(a) the optimal design and functioning of parallel computational systems, (b) the optimal recognition of optical and acoustic fields in synthesizing an optimal dynamic analyzer, and (c) the modeling of nonlinear transfer processes in the component base of a computer.

We discuss new mathematical methods that can be applied in solving specific problems arising in the construction of mathematical models for handling the above-mentioned three topics. Although various countries have developed devices and technological processes for creating new generations of computers, there is still no general theoretical approach. In this respect the present collection fills an important gap in the literature on the subject.

All results set forth in this collection are new and obtained only recently. Here we give a brief survey.

The article written by S. M. Avdoshin, V. V. Belov, V. P. Maslov, and A. M. Chebotarev is devoted to constructing a theory of the optimization problems that emerge in the development of the architecture, the organization of parallel computations, and the design of flexible manufacturing systems for homogeneous multiprocessor computational systems. The following concept lies at the base of the suggested approach: all the optimization problems considered here are linear in a space of functions with values in semirings. Depending on the choice of the semimodule, for instance, the Hamilton-Jacobi equation and the Bellman equation prove to be linear in the new sense. For these equations analogs of the Duhamel principle and the Fredholm alternatives prove valid. This concept leads to a new definition of an integral corresponding to the semigroup operation of the "sum" type, the concept of a measure additive in this new sense, and an analog of the scalar product (say, $(\varphi_1, \varphi_2) = \min\limits_x [\varphi_1(x) + \varphi_2(x)]$ or $(\varphi_1, \varphi_2) = \max\limits_x [\min\limits_x \varphi_1(x), \varphi_2(x)])$, which makes it possible to go over to adjoint operators and define functions "generalized" in the new sense. On the basis of the "linearity" concept and the notion of generalized convergence in optimization problems dealing with homogeneous computational systems, we consider the passage to the limit in a natural large parameter proportional to the number of elementary processors in the computational system. For a broad class of problems the solutions to the limiting equations can be set up on the basis of Pontryagin's maximum principle.

The article by V. P. Belavkin and V. P. Maslov has a direct bearing on the problem of mathematical synthesis of an optimal

dynamic analyzer, a device intended for automatic sound recognition. As is known, establishing a verbal link between man and computer is one of the key problems in the design of fifth-generation computational systems. The article provides a systematic exposition of the wave theory of representations and measurements, the theory that is based on similarities with quantum mechanics and used to solve problems of detection, separation, identification, and estimation of the parameters of acoustic and visual images within the framework of the noncommutative theory of wave hypothesis testing. The idea of applying quantum mechanics to the problem of recognizing wave images emerged at the beginning of the 1970s, when a seminar devoted to quantum mechanics and image recognition was opened in the Physics Department of Moscow State University under the direction of Yu. P. Pyt'ev and this author.

The article by V. G. Danilov, V. P. Maslov, and K. A. Volosov is directly related to the key issue of creating the component base of computers, namely, the calculation and design of new technological methods for the various stages of designing integrated circuits and other computer elements. Mathematical modeling in this case is a preliminary stage. Most of the modeling problems can be reduced to a quasilinear parabolic equation or a system of such equations.

The article suggests new methods for building asymptotic equations to quasilinear parabolic equations. From the mathematical view the class of problems considered is characterized by two effects: localization of a perturbation and the finiteness of the speed with which the perturbation travels. In other words, the support of the solution is a compact set or a semibounded set, and the boundary of the support propagates with a certain speed. At the support boundary the solution undergoes a weak discontinuity; hence, along with the problem of constructing the asymptotics in a small parameter there emerges the problem of the propagation of the singularity (the weak discontinuity). The theory developed in the article is applied to calculating such processes as diffusion, heat conduction, turbulent filtration, adsorption (desorption), epitaxy, and film flow. Application of the findings to the various stages of the technological processes reduces designing time and production costs.

The abundance of basically new material, that is, new methods, notions, definitions, etc., must have posed certain difficulties in preparing the manuscript for print. For this reason I would like to express my sincere gratitude to Mir Publishers for undertaking to introduce the foreign reader to the achievements in this field of knowledge. In particular, I would like to thank the staff of the mathematics editorial office for preparing the Russian version of the manuscript for translation and the English physics and mathematics editorial office for the expert translation.

January 1988 *Academician V. P. Maslov*

1 Design of Computational Media: Mathematical Aspects

S. M. Avdoshin, V. V. Belov, V. P. Maslov,
and *A. M. Chebotarev*

1.0 A Brief Survey

In the present article we aim at a solution of certain problems associated with the architecture and analysis of parallel programs and flexible manufacturing systems (FMS) for homogeneous multiprocessor computational systems (CS), which are characteristic of fifth-generation computers.

These problems constitute examples of optimization problems containing a natural large parameter proportional to N, the number of elementary processors in the CS. As a rule, the complexity of the solution algorithms for these problems increases rapidly with N (at least like N^2). This brings us to the problem of the limiting transition as $N \to \infty$, that is, a limit problem whose solution does not depend on N and approximates the solution of the initial problem all the better as N increases. In some important specific cases this limit problem can be associated with the Bellman equation [1.1]. As a rule, however, even for smooth initial data the limit equation has no differentiable solutions (except for a small number of extremely special problems). Hence, the classical statement of the Cauchy problem for this equation usually has no meaning. More than that, the Bellman equation does not even have generalized solutions in the usual sense. Hence only Pontryagin's maximum principle applied to such cases has a clearly defined mathematical meaning. This principle has been used in solving the corresponding optimization problems [1.2].

The general approach suggested in this paper to solving optimization problems related to multiprocessor computers can also be applied to optimization problems of an entirely different nature. This approach is based on the fact that all optimization problems considered here are "linear" in function spaces whose elements have values in certain semi-rings. Here is what this means. Let us consider a function space in which the common operations of addition and multiplication by numbers are replaced with other semi-group operations, $\oplus$ and $\odot$, related through the distributivity law. For instance, instead of the sum of two functions we take their supremum, and instead of the product of a function by a number we take the infimum. The linearity of equations in such spaces means that, if $y_1(x)$ and

$y_2(x)$, $x \in X$, are solutions, then $\sup\limits_{x \in X}(y_1(x),\ y_2(x))$ and $\inf\limits_{x \in X}(y_1(x), \lambda)$ or $\inf\limits_{x \in X}(y_2(x), \lambda)$, $\lambda = \text{const}$, are also solutions. Next we introduce the concept of an "integral" corresponding to a semi-group operation of the "sum" type, the concept of measure that is additive in this new sense, and an analog of the scalar product, which makes it possible to introduce conjugate operators and define functions that are "generalized" in the new sense. For example, the scalar product in a space of functions with values in a semi-ring A where the sum is replaced with min and the product with the common sum, $+$, has the form

$$\langle \varphi_1,\ \varphi_2 \rangle = \min_{x \in X}(\varphi_1(x) + \varphi_2(x)) \overset{\text{def}}{=} \int_{\oplus} \varphi_1(x) \odot \varphi_2(x)\, dx.$$

In this space, for the Hamilton-Jacobi equation

$$\frac{\partial u}{\partial t} + H\left(x,\ \frac{\partial u}{\partial t},\ t\right) = 0 \tag{$*$}$$

we have the superposition principle for solutions, that is, if y_1 and y_2 are solutions, then $\lambda_1 \odot y_1 \oplus \lambda_2 \odot y_2$, with $\lambda_i = \text{const}$ $(i = 1, 2)$, are also solutions. This leads to a formula that represents a solution of the equation in terms of a source, or

$$\begin{aligned} u(x,\ t) &= \int_{\oplus} k(x,\ \zeta,\ t) \odot u_0(\zeta)\, d\zeta \\ &= \min_{\zeta \in X}(k(x,\ \zeta,\ t) + u_0(\zeta)), \end{aligned} \tag{1.0.1}$$

where $k(x,\ \zeta,\ 0) = \delta(x - \zeta)$, and $\delta(x - \zeta)$ is understood to be the functional $\min\limits_{\xi \in X}(\delta(x - \zeta) + \varphi(\zeta)) = \varphi(x)$, say, $\delta(x - \zeta) = \lim\limits_{\varepsilon \to 0}[(x - \zeta)^2/\varepsilon]$. It is easy to see that $k(x,\ \zeta,\ t) = \min \int \mathcal{L}\, dt$, where $\mathcal{L}$ is the Lagrangian, and formula (1.0.1) proves to be the well-known representation of a solution to the Hamilton-Jacobi equation $(*)$ in the small in terms of a generating function.

The "Fourier transform" in a space of functions with the values in a semi-ring A is the eigenfunction expansion of the translation (or shift) operator T_Δ, that is, $T_\Delta \varphi(x) = \varphi(x + \Delta)$; the eigenfunctions $\psi_\mu(x)$ of T_Δ have the form μx: $T_\Delta \mu x = \mu(x + \Delta x) = \mu\Delta + \mu x = \mu\Delta \odot \mu x$, and the corresponding eigenvalues are $\mu\Delta$. The "Fourier transform" of a function $\varphi(x)$ has the form $\int_{+} \psi_\mu(x) \odot \varphi(x)\, dx = \min\limits_{x \in X}(\mu x + \varphi(x))$ and in the case at hand coincides with the Legendre transform, which is "linear" in this space. It has been established that if $H(p, x, t)$ is a function homogeneous of degree one in p,

then the Cauchy problem for equation (*) is also "linear" in the space of functions with the values in the semi-ring A: $\oplus = \min$, $\odot = \max$. The Cauchy problem for the Bellman equation also proves to be "linear" in appropriate function spaces of functions with values in a semi-ring A.

The general differential equation in spaces of functions of a continuous argument that generalizes both the Bellman equation and the Hamilton-Jacobi equation is an equation for which the resolving operator is linear in function spaces with values in the appropriate semi-rings and whose solution is generalized in the above-mentioned new sense (that is, is a "linear" continuous functional with respect to the new "scalar product"). We will call this equation the generalized Hamilton-Jacobi equation and in the discrete case the generalized Bellman equation. For such equations there exists an analog of Fredholm alternative theorems. For a broad class of problems the solutions of these equations can be constructed using Pontryagin's maximum principle as a basis.

The concept described above makes it possible to determine the limiting values when $N \to \infty$ and overcome the difficulty that arises from the fact that usually the solutions of such problems assume only two values, 0 and 1. For the sake of comparison we first turn to the linear case, where the discrete problem converges to a continuous one.

Example 1. Suppose a discrete problem is described by the difference scheme

$$a_n^{k+1} = La_n^k, \quad n \in \mathbb{Z}, \quad k \in \mathbb{Z}_+, \tag{1.0.2}$$

where L is a linear difference operator with constant coefficients on an integer lattice, with the initial value a_n^0 a nonzero constant (say, c) for $n \geqslant 0$ and zero for $n < 0$. This problem has no limit in the ordinary sense of the word as $k \to \infty$. Nevertheless, the concept of a generalized solution introduced by S. L. Sobolev makes it possible to find the weak limit of the solution to this problem. To this end we take a family of functions of continuous independent variables, $v_h(t, x)$, $t \in [0, T]$, $T = \text{const}$, $x \in R^1$, that depends on parameter $h \in (0, 1]$ and is such that

$$v_h(kh, nh) = a_n^k.$$

With the initial problem (1.0.2) we associate the problem for $v_h(t, x)$:

$$v_h(t + h, x) = \widetilde{L}_h v_h(t, x), \quad v_h|_{t=0} = v^0(x), \quad t = kh. \tag{1.0.2'}$$

Here $\widetilde{L}_h$ is the natural continuation of L on functions of a continuous independent variable. For example, if $La_n^k = \sum_i c_i a_{n+i}^k$, then

$\widetilde{L}_h v_h(t, x) = \sum_i c_i v_h(t, x + ih)$. The condition $k \to \infty$ is equivalent to $h \to 0$, since $kh \leqslant T$. Let us assume that on smooth functions the operator $(\widetilde{L}_h)^l$ converges to the operator e^{tL_0} as $l \to \infty$, $lh \to t$, where L_0 is a linear differential operator that, as $h \to 0$, is approximate on smooth functions by the operator $[\widetilde{L}_h - 1]/h$. Then, if the initial value $v^0(x)$ is a smooth function, the family of functions $v_h(lh, x)$ converges (in $C(R^1_x)$) to the solution $v(t, x)$ of the differential equation

$$\frac{\partial v}{\partial t} = L_0 v, \quad v|_{t=0} = v^0(x). \tag{1.0.3}$$

If $v^{(0)}(x)$ is a discontinuous function, the solution to the difference problem converges to a solution of the differential equation in the sense of generalized functions. Indeed, for any smooth finite function φ we have

$$\begin{aligned}(v_h(lh, x), \varphi) &= ((\widetilde{L}_h)^l v^0, \varphi) = (v^0, (\widetilde{L}^*_h)^l \varphi) \\ \xrightarrow[h=0]{} (v_0, e^{tL^*_0}\varphi) &\overset{\text{def}}{=} (e^{tL_0}v^0, \varphi) = (v(t, x), \varphi) \\ &\overset{\text{def}}{=} \int \varphi(x)\, v(t, x)\, dx,\end{aligned}$$

where $v(t, x)$ is a generalized solution to problem (1.0.3).

We have therefore found that the weak limit of the solution to a difference problem is a generalized solution to the respective limiting equation, to which the initial difference equation converges only on smooth functions.

Let us now study the analogy between the example just considered and the solution to the respective discrete optimization problem.

Example 2. Consider the process $\{a^k, k = 0, 1, \ldots\}$ with a discrete space of states $\mathbb{Z} \times \mathbb{Z}$, $a^k\colon \mathbb{Z} \times \mathbb{Z} \to R^1$, and satisfying the Bellman equation

$$\begin{aligned}&a^{k+1}(m, n) = \min\{a^k(m-1, n), a^k(m, n-1)\}, \\ &n, m \in \mathbb{Z}, \quad k = 0, 1, \ldots,\end{aligned} \tag{1.0.4}$$

and the initial data of the form

$$a^0(m, n) = \begin{cases} 0 \text{ if } n = 0, m \geqslant 0 \text{ or } m = 0, n \geqslant 0, \\ +\infty \text{ otherwise.}\end{cases} \tag{1.0.5}$$

Note that this equation is linear in the space of functions with discrete arguments with values in the semi-ring $A = (R^1 \cup \{\pm\infty\}, \oplus = \min, \odot = +)$, where $\mathbb{1} = 0$ and $\mathbb{0} = +\infty$. We may rewrite it in a form quite similar to the one discussed in Example 1:

$$a^{k+1}(m, n) = L_\nabla a^k(m, n), \tag{1.0.6}$$

where the "linear" operator L_V is given by the formula

$$L_V a^k(m, n) = \bigoplus_{(i,j)\in V} c_{ij} \odot a^k(m-i, n-j),$$

$$c_{ij} = \mathbb{1} = 0, \ V = \{(0, +1), (+1, 0)\},$$

and the initial value is

$$a^0(m, n) = \begin{cases} \mathbb{1} \text{ if } n=0, \ m \geqslant 0 \text{ or } m=0, \ n \geqslant 0, \\ \mathbb{0} \text{ otherwise.} \end{cases} \quad (1.0.7)$$

Just as in the linear case, this problem has no limit if we send k to ∞. Nevertheless, the concept of "generalized" solutions makes it possible, as in the linear case, to obtain the weak limit of problem (1.0.6), (1.0.7). With problem (1.0.6), (1.0.7) we associate the following problem for functions of continuous arguments $(x, y) \in R^2 \cup \{\pm\infty\}$, $t \in [0, T]$, $T = \text{const} > 0$:

$$u_h(t+h, x, y) = \tilde{L}_{h,V} u_h(t, x, y), \ t = kh, \ k \in \mathbb{Z}, \quad (1.0.6')$$

$$u_h(0, x, y) = u^0(x, y) = \begin{cases} \mathbb{1} \text{ if } x=0, \ y \geqslant 0 \text{ or } y=0, \ x \geqslant 0, \\ \mathbb{0} \text{ if } x \neq 0 \text{ or } y \neq 0, \end{cases} \quad (1.0.7')$$

in such a manner that $u_h(kh, mh, nh) = a^k(m, n)$.

The operator $\tilde{L}_{h,V}$ in the given case acts according to the formula

$$\tilde{L}_{h,V} u_h(t, x, y) = \min(u_h(t, x-h, y), \ u_h(t, x, y-h)). \quad (1.0.8)$$

At $t = kh$ the solution to problem (1.0.6′), (1.0.7′) assumes the form

$$u_h(t, x, y) = (\tilde{L}_{h,V})^k u^0(x, y).$$

Allowing for (1.0.8), we can calculate the right-hand side of this equation explicitly:

$$u_h(t, x, y) = \min_{\substack{\zeta_1+\zeta_2=k \\ \zeta_i \in \mathbb{Z}_+ \\ i=1,2}} \{u^0(x-h\zeta_1, y-h\zeta_2)\}.$$

Let us introduce the "scalar product" for A-valued functions assuming that

$$\langle \varphi, \psi \rangle_\oplus = \int_\oplus \varphi(x, y) \odot \psi(x, y)\, dx\, dy \stackrel{\text{def}}{=} \inf_{x,y} \{\varphi(x, y) + \psi(x, y). \quad (1.0.9)$$

Now we wish to calculate the weak limit, as $h \to 0$, of the solution to problem (1.0.6′), (1.0.7′) on smooth functions with respect to the

"scalar product" introduced above (this, as noted earlier, is equivalent to finding the limit as k tends to ∞). Suppose $k \to \infty$, $kh \to t_0$, $t_0 \in [0, T]$. Then for every smooth function $\varphi(x, y)$ we have

$$\langle u_h(t_0, x, y), \varphi(x, y)\rangle_\oplus = \langle (\widetilde{L}_{h,V})^k u^0(x, y), \varphi(x, y)\rangle_\oplus$$
$$= \langle u^0(x, y), (\widetilde{L}^*_{h,V})^k \varphi(x, y)\rangle_\oplus,$$

where operator $\widetilde{L}^*_{h,V}$ is the conjugate of $\widetilde{L}_{h,V}$ with respect to the scalar product $\langle\ ,\ \rangle_\oplus$ introduced above.

It can easily be verified that

$$(\widetilde{L}^*_{h,V})^k \varphi(x, y) = \min_{\substack{\zeta_1+\zeta_2=k \\ \zeta_i \in \mathbb{Z}_+}} \{\varphi(x + h\zeta_1, y + h\zeta_2)\}.$$

Hence

$$\langle u_h(t_0, x, y), \varphi(x, y)\rangle_\oplus$$
$$= \langle u^0(x, y), \min_{\substack{\zeta_1+\zeta_2=k \\ \zeta_i \in \mathbb{Z}_+}} \{\varphi(x + h\zeta_1, y = h\zeta_2)\}\rangle_\oplus$$

We denote by $u_0(t_0, x, y)$ the weak limit of the solution to problem (1.0.6′), (1.0.7′): $u_0(t_0, x, y) = s - \lim_{h\to 0} u_h(t_0, x, y)$, where s is a fixed vector function whose meaning will be defined later on. Sending h to 0 and kh to t_0 in the last equation, we get

$$\langle u_0(t_0, x, y), \varphi(x, y)\rangle_\oplus$$
$$= \lim_{h\to 0} \langle u_h(t_0, x, y), \varphi(x, y)\rangle_\oplus$$
$$= \langle u^0(x, y), \min_{\substack{\eta_1+\eta_2=t_0 \\ \eta_i \in R^1_+}} \{\varphi(x + \eta_1, y + \eta_2)\}\rangle. \qquad (1.0.10)$$

We will call the generalized function $u_0(t_0, x, y)$ defined by (1.0.10) a generalized solution to the limiting generalized Hamilton-Jacobi equation to which Eq. (1.0.6′) converges on smooth functions. This limiting equation has the form

$$\frac{\partial u}{\partial t} = \min\left\{-\frac{\partial u}{\partial x}, -\frac{\partial u}{\partial y}\right\}.$$

The solution to this equation can be obtained by employing Pontryagin's maximum principle [1.2, 1.3].

In contrast to the linear case, the statement of smooth initial data for discrete optimization problems has no meaning, as a rule. For this reason a study of the limiting equation of an optimization problem is justified only in constructing generalized solutions to this equation in the above sense.

The example of the discrete optimization problem (1.0.4), (1.0.5) is closely related to an analysis of the activity of homogeneous multiprocessor computational systems (see Sec. 1.2). When the number of processors, N, in such a system grows, that is, $N \to \infty$ ($h \sim 1/N \to 0$), the support of the generalized solution (1.0.10) to the limiting Hamilton-Jacobi equation determines, at each moment $t > 0$, the set of processors carrying out calculations at time t.

The suggested approach enables considering generalized solutions for general optimization problems, too. But here we will give a brief description of the properties of discrete optimization problems that arise when the operation of homogeneous computational systems is analyzed.

It appears that all such problems can be studied using solutions (generalized solutions in the limiting case of $N \to \infty$) to the generalized Hamilton-Jacobi equation in the space of functions with values in semi-rings. It has also been found that in problems related to the architecture of multiprocessor homogeneous CS, the range of the sought functions has the structure of a crystal lattice with certain symmetry properties. For instance, in the simplest case of a matrix processor, a draft of which was proposed in 1982 by a group of US scientists [1.4-1.7], the common Bravais lattice [1.8] serves as such a range.

For optimization problems connected with the estimation of the effectiveness of parallel programs, a discrete lattice with nontrivial symmetry groups (a nonempty set of nonelementary translations) serves as a natural range of independent variables of the functions involved in the problems. The symmetry of such lattices is uniquely determined by the text of the program, while the execution time of the program operators is determined by the values of the coefficients of the appropriate system of generalized Hamilton-Jacobi equations (systems of discrete generalized Bellman equations).

In optimization problems related to the operation of CS, these equations are usually nonhomogeneous steady-state equations (the right-hand side of the equations describes the interaction of the set of processors in a CS with the external memory of the system). The solution to these equations is found as the limit, as $N \to \infty$, of the solutions of appropriate evolutionary nonhomogeneous equations. Solving the latter can be reduced to solving homogeneous equations, using an analog of Duhamel's principle.

The difference between this case and the common linear case lies in the following: although for a steady-state problem there is no limiting generalized Hamilton-Jacobi equation, in the limit the corresponding nonstationary problem can be reduced to the evolutionary generalized Hamilton-Jacobi equation. This makes it possible, by means of Duhamel's principle, to write the limiting problem

in terms of generalized solutions of a certain Cauchy problem for the generalized Hamilton-Jacobi equation. We call such a problem a stabilization one. Thus, a generalized solution of a stabilization Cauchy problem is the limit (in the new sense of the word), as $t \to \infty$, of the solution to the Cauchy problem for the generalized Hamilton-Jacobi equation.

Let us illustrate the aforesaid with two examples. In the first example we will consider a nonhomogeneous steady-state scalar equation on a simple one-dimensional lattice, so as to demonstrate how Duhamel's principle can be employed. In the second example we will study an optimization problem on a two-dimensional discrete lattice with a nontrivial symmetry group.

Example 3. Let us consider the simplest one-dimensional "tracing" problem, the problem of finding the shortest route [1.9-1.11] on a discrete lattice $\Omega_\varepsilon = \{x \quad x_n = n\varepsilon, n = 0, \pm 1, \ldots\}$ with spacing ε, where ε is a positive parameter. The following relation exists for the length $s_\varepsilon(n)$ of the shortest route at point n:

$$s_\varepsilon(n) = \min\{s_\varepsilon(n-1) + \varepsilon c_1,\ s_\varepsilon(n+2) + \varepsilon c_2,\ \mathcal{F}_\varepsilon(n)\}, \quad n \in \mathbb{Z}, \tag{1.0.11}$$

where c_1 and c_2 are constants, and $\mathcal{F}_\varepsilon(n) = g(n\varepsilon)$, with $g(x)$, $x \in R^1$, a continuous function bounded below.

Problem (1.0.11) is a steady-state problem with a right-hand side equal to $\mathcal{F}_\varepsilon(n)$ and is linear in the space of functions with values in the semi-ring

$$A = \{R^1 \cup \{\pm\infty\}, \quad \oplus = \min, \quad \odot = +\},$$

where $\mathbb{0} = +\infty$, and $\mathbb{1} = 0$. We rewrite it in the form

$$s_\varepsilon(n) = L_\varepsilon s_\varepsilon(n) \oplus \mathcal{F}_\varepsilon(n)$$

where operator L_ε acts according to the rule

$$L_\varepsilon s_\varepsilon(n) = \bigoplus_{v \in V} c_\varepsilon(v) \odot s_\varepsilon(n - v),$$

with $V = \{v_1 = 1,\ v_2 = -2\}$, $c_\varepsilon(v_1) = \varepsilon c_1$, and $c_\varepsilon(v_2) = \varepsilon c_2$.

With this problem we associate the following problem for the family of functions of a continuous variable $u_\varepsilon(x)$, $x \in R^1$, $\varepsilon \in (0, 1]$:

$$u_\varepsilon(x) = \tilde{L}_\varepsilon u_\varepsilon(x) \oplus g(x), \tag{1.0.11'}$$

where operator $\tilde{L}_\varepsilon$ is defined as follows:

$$\tilde{L}_\varepsilon u_\varepsilon(x) = \min\{u_\varepsilon(x - \varepsilon) + \varepsilon c_1,\ u_\varepsilon(x + 2\varepsilon) + \varepsilon c_2\}.$$

Obviously, $u_\varepsilon(n\varepsilon)$ is the solution to the initial discrete problem (1.0.11). The solution to problem (1.0.11′) is the limit, as $t \to \infty$, of the solution to the evolutionary nonhomogeneous equation

$$f_\varepsilon(t + \varepsilon, x) = \tilde{L}_\varepsilon f(t, x) \oplus g(x), \quad t = k\varepsilon, \quad k = 0, 1, \ldots, \tag{1.0.12}$$

with the initial data

$$f_\varepsilon(0, x) = \mathbb{0} = +\infty.$$

Just as in the linear case, by virtue of Duhamel's principle (see Sec. 1.1), solution $f_\varepsilon(t, x)$ is the integral (in the sense of $\oplus$) of solution $W_\varepsilon(t, \tau, x)$ of the Cauchy problem

$$W_\varepsilon(t+\varepsilon, \tau, x) = \widetilde{L}_\varepsilon W_\varepsilon(t, \tau, x), \quad t \geqslant \tau,$$
$$W_\varepsilon(t, \tau, x)\,|_{t=\tau} = g(x).$$

In the case at hand,

$$f_\varepsilon(t, x) = \int_{\oplus}^{[0,t]} W_\varepsilon(t, \tau, x)\, d\tau$$
$$\stackrel{\text{def}}{=} \min_{0 \leqslant \tau \leqslant t} W_\varepsilon(t, \tau, x). \tag{1.0.13}$$

Similar to the solution to problem (1.0.6′), (1.0.7′) in Example 2, we can calculate

$$W_\varepsilon(t, \tau, x) = \min_{\substack{\zeta_1+\zeta_2=k-l \\ \zeta_i \in \mathbb{Z}_+}} \{\varepsilon c_1\zeta_1 + \varepsilon c_2\zeta_2 + g(x - \varepsilon\zeta_1 + 2\varepsilon\zeta_2)\}, \tag{1.0.14}$$

where $t = k\varepsilon$ and $\tau = l\varepsilon$, $k \geqslant l$, k, $l \in \mathbb{Z}_+$.

Hence, from (1.0.13) and (1.0.14) it follows that

$$f_\varepsilon(t, x) \stackrel{\text{def}}{=} R_\varepsilon(t)\, g(x)$$
$$= \min_{0 \leqslant \varepsilon\alpha \leqslant t} \min_{\substack{\varepsilon\zeta_1+\varepsilon\zeta_2=\varepsilon\alpha \\ \zeta_i \in \mathbb{Z}_+}} \{\varepsilon c_1\zeta_1 + \varepsilon c_2\zeta_2 + g(x - \varepsilon\zeta_1 + 2\varepsilon\zeta_2)\} \tag{1.0.15}$$

where we have introduced the notation $k - l = \alpha$, $\alpha \in \mathbb{Z}_+$. If we now send t to ∞, we find the solution $u_\varepsilon(x)$ to Eq. (1.0.11′):

$$u_\varepsilon(x) = \lim_{t\to\infty} f_\varepsilon(t, x) = \inf_{\substack{\varepsilon\zeta_i \geqslant 0 \\ \zeta_i \in \mathbb{Z}_+}} \{\varepsilon c_1\zeta_1 + \varepsilon c_2\zeta_2$$
$$+ g(x - \varepsilon\zeta_1 + 2\varepsilon\zeta_2)\}. \tag{1.0.16}$$

We denote by $u(x)$ the strong limit of the solution to problem (1.0.11′):

$$u(x) = \lim_{\varepsilon\to 0} u_\varepsilon(x), \quad x \in R^1.$$

Sending ε to 0 in this equation, we obviously arrive at

$$u(x) = \inf_{\eta_i \in R^1_+} \{c_1\eta_1 + c_2\eta_2 + g(x - \eta_1 + 2\eta_2)\}.$$

Now suppose that $g(x)$ in Eq. (1.0.11′) is a discontinuous function that assumes only two values, say

$$g(x)=\begin{cases}\mathbb{1} & \text{if } x \geqslant 0,\\ \mathbb{O} & \text{if } x<0.\end{cases}$$

The concept of generalized solutions (in the new sense of the word) makes it possible to calculate in this case the weak limit of problem (1.0.11′) with respect to the "scalar product" for A-valued functions (for the meaning of the "scalar product" see Eq. (1.0.9)). This limit is the generalized solution of the stabilization Cauchy problem for the limiting generalized Hamilton-Jacobi equation, to which Eq. (1.0.12) is reduced on smooth functions. The stabilization Cauchy problem for the limiting equation has the following form:

$$\frac{\partial u}{\partial t}=\min\left\{c_1-\frac{\partial u}{\partial x},\ c_2+2\frac{\partial u}{\partial x}\right\}, \tag{1.0.17}$$
$$u|_{t=0}=g(x),\quad s-\lim_{t\to\infty} u(t,\ x)=W(x).$$

Let us find the generalized solution $W(x)$ to problem (1.0.17) using the explicit form of the resolving operator $R_\varepsilon(t)$ of problem (1.0.12) defined in (1.0.15). As in Example 2, it is easy to verify that $R_\varepsilon^*(t)$ is the conjugate of $R_\varepsilon(t)$ with respect to the scalar product $\langle\ ,\ \rangle_\oplus$ and operates according to the rule

$$R_\varepsilon^*(t)\,\psi(x)=\min_{\substack{0\leqslant k-l\leqslant k\\ \zeta_1+\zeta_2=k-l\\ \zeta_i\in\mathbb{Z}_+}}\{\varepsilon c_1\zeta_1+\varepsilon c_2\zeta_2+\psi(x+\varepsilon\zeta_1-2\varepsilon\zeta_2)\}.$$

Combining this with (1.0.15) and (1.0.16), we find that for every smooth finite function $\varphi(x)$,

$$\langle W(x),\varphi(x)\rangle_\oplus \overset{\text{def}}{=} \lim_{\varepsilon\to 0}\langle u_\varepsilon(x),\ \varphi(x)\rangle_\oplus = \lim_{\varepsilon\to 0}\lim_{t\to\infty}\langle R_\varepsilon(t)\,g(x),\ \varphi(x)\rangle_\oplus$$
$$=\lim_{\varepsilon\to 0}\lim_{t\to\infty}\langle g(x),\ R_\varepsilon^*(t)\,\varphi(x)\rangle_\oplus$$
$$=\lim_{\varepsilon\to 0}\lim_{t\to\infty}\langle g(x),\ \min_{0\leqslant\varepsilon\alpha\leqslant t}\ \min_{\varepsilon\zeta_1+\varepsilon\zeta_2=\varepsilon\alpha}\{\varepsilon c_1\zeta_1+\varepsilon c_2\zeta_2+\varphi(x+\varepsilon\zeta_1-2\varepsilon\zeta_2)\rangle_\oplus$$
$$=\langle g(x),\ \inf_{\eta_i\in R^1_+}\{c_1\eta_1+c_2\eta_2+\varphi(x+\eta_1-2\eta_2)\}\rangle_\oplus. \tag{1.0.18}$$

The generalized function $W(x)$ defined by (1.0.18) is the weak limit of the initial optimization steady-state problem (1.0.11).

Example 4. Let us now take a discrete optimization problem related to the choice (design) of the architecture of a homogeneous multiprocessor CS optimal from the viewpoint of effectiveness of parallel programs in such CS (see Section 1.2.3). For instance, for a parallel program $P_{A\cdot n}$ for multiplying two square matrices this problem

can be reduced to solving, on the discrete lattice

$\Omega = \Omega_1 \cup \Omega_2 = \{r = (p, q), p = n_1 + 1/2, q = n_2 + 1/2\}$
$\cup \{r = (p, q), p = n_1 + 1/3, q = n_2 + 1/4\}, \quad n_i \in \mathbb{Z}, \quad i = 1, 2,$

the equation of the following form:

$$u(n_1 + 1/2, n_2 + 1/2) = \max\{\mathcal{F}_1(n_1, n_2), u(n_1 + 1/2 - 1, n_2 + 1/2) + t_0, u(n_1 + 1/3 - 1, n_2 + 1/4 - 1) + t_0\} \quad (1.0.19)$$

at a point $r \in \Omega_1$, and

$$u(n_1 + 1/3, n_2 + 1/4) = \max\{\mathcal{F}_2(n_1, n_2), u(n_1 + 1/2, n_2 + 1/2) + t_0, u(n_1 + 1/3, n_2 + 1/4 - 1) + t_0\}$$

at a point $r \in \Omega_2$. Here (n_1, n_2) are the coordinates of the elementary processors of which the CS consists, $t_0 = \text{const} > 0$ is the time of execution of interprocessor data exchange operations in the $P_{A\times B}$ program, and the given functions $\mathcal{F}_i(n_1, n_2) > 0$, $i = 1,2$ are determined by the "loading" times of local subprograms (see p. 92). The value of the function $u(r)$, $r \in \Omega$, is the completion time of execution of the $P_{A\times B}$ program on the $n = (n_1, n_2)$ processor. Note that this equation is linear in the space of functions of discrete arguments with values in the semi-ring $A = (R^1 \cup \{\pm\infty\}, \oplus = \max, \odot = +)$, where $\mathbb{0} = -\infty$ and $\mathbb{1} = 0$, while the range of definition for the sought functions is the lattice Ω in R^2 with an additive group of elementary translations generated by the shifts T_{α_1} and T_{α_2} by vectors $\alpha_1 = (1, 0)$ and $\alpha_2 = (0, 1)$ and with a set of nonelementary translations generated by shifts by the basis vectors $\beta_1 = (1/2, 1/2)$ and $\beta_2 = (1/3, 1/4)$.

Let us find the weak limit of Eq. (1.0.19). With Eq. (1.0.19) we associate an equation for the family of functions of continuous arguments, $v_h(x)$, $x \in R^2$, the family depending on parameter $h \in (0\ 1]$, or

$$v_h(x) = \tilde{L}_h(x)\, v_h(x) \oplus \Phi_h(x), \quad (1.0.19')$$

in such a manner that $v_h(nh) = u(n)$, $n = (n_1, n_2)$,

$$\Phi_h(nh) = \begin{cases} \Phi_h^1(nh) = \mathcal{F}_1(n) & \text{if } n + \beta_1 \in \Omega_1, \\ \Phi_h^2(nh) = \mathcal{F}_2(n) & \text{if } n + \beta_2 \in \Omega_2 \end{cases}$$

where $\Phi_h^i(x)$, $i = 1, 2$, are families of continuous functions, with $\Phi_h^i(x) \to \Phi_0^i(x)$ as $h \to 0$, and $\beta_1 = (1/2, 1/2)$, $\beta_2 = (1/3, 1/4)$. The linear operator $\tilde{L}_h(x)$ depends on the point x where the result of its action on function v_h is calculated and is determined by the following formulas:

$$(\tilde{L}_h(x)\, v_h(x) = \max\{\Phi_h(x), v_h(x + h\beta_1 - hv_2) + ht_0, v_h(x + h\beta_2 - hv_4) + ht_0\}$$

for $x \in \Omega_1$, and

$$(\widetilde{L}_h(x) v_h)(x) = \max\{\Phi_h(x),\ v_h(x + h\beta_1) + ht_0,\ v_h(x + h\beta_2 - hv_3) + ht_0\} \tag{1.0.20}$$

for $x \in \Omega_2$, where $v_2 = (1, 0)$, $v_3 = (0, 1)$, and $v_4 = (1, 1)$.

Since the scalar operator $\widetilde{L}_h(x)$ is not a regular function of x and h, even on smooth functions it possesses no limit as $h \to 0$, and, hence, the corresponding stabilization Cauchy problem has no limit either. A similar situation emerges in the linear case when we wish to study the vibrations of atoms in a crystal lattice with a nonempty set of nonelementary translations [1.12].

To overcome this difficulty, instead of the scalar equation (1.0.20) we consider an equivalent system of equations for a family of A-valued vector functions s_h: $R^2 \to A \times A$. We assume that $s_h = (s_h^1, s_h^2)^{\mathrm{T}}$, with $s_h^1(x) = v_h(x + h\beta_1)$, $s_h^2(x) = v_h(x + h\beta_2)$, and the superscript T standing for "transposition", and define A-valued 2-by-2 matrices thus:

$$\sigma_1 = \begin{bmatrix} \mathbb{0} & \mathbb{0} \\ \mathbb{1} & \mathbb{0} \end{bmatrix},\ \sigma_2 = \begin{bmatrix} \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{0} \end{bmatrix},\ \sigma_3 = \begin{bmatrix} \mathbb{0} & \mathbb{0} \\ \mathbb{0} & \mathbb{1} \end{bmatrix},\ \sigma_4 = \begin{bmatrix} \mathbb{0} & \mathbb{1} \\ \mathbb{0} & \mathbb{0} \end{bmatrix}. \tag{1.0.21}$$

The system of equations for the family s_h corresponding to (1.0.20) assumes the form

$$s_h = \hat{L}_{h,V} s_h \oplus \widetilde{\Phi}_h,\quad \widetilde{\Phi}_h = (\Phi_h^1, \Phi_h^2)^{\mathrm{T}}, \tag{1.0.22}$$

where the matrix operator $\hat{L}_{h,V}$ is defined thus:

$$(\hat{L}_{h,V} S_h)(x) = (ht_0) \odot \left(\bigoplus_{v \in V} \hat{\sigma}(v)\, s_h(x - vh) \right). \tag{1.0.23}$$

Here V is the set of vectors $\{v_1 = (0, 0),\ v_2 = (1, 0),\ v_3 = (0,1),\ v_4 = (1, 1)\}$, and $\hat{\sigma}(v_i) \overset{\mathrm{def}}{=} \sigma_i$, $i = 1, 2, 3, 4$.

We will denote the "scalar product" for A-valued vector functions by $\langle \cdot, \cdot \rangle_{A^2}$, assuming all along that

$$\langle \varphi, \psi \rangle_{\oplus} = \int_{\oplus} \langle \varphi(x), \psi(x) \rangle_{A^2}\, dx \overset{\mathrm{def}}{=} \bigoplus_x \bigoplus_{j=1,2} \varphi^i(x) \odot \psi^i(x) = \max_x \max_{j=1,2} \{\varphi^i(x) + \psi^i(x)\}.$$

We can easily see that $\hat{L}_{h,V}^*$, the conjugate of operator (1.0.23) with respect to the scalar product we have just introduced, is defined thus:

$$(\hat{L}_{h,V}^* \psi)(x) = (ht_0) \odot \left(\bigoplus_{v \in V} \hat{\sigma}(v)\, \psi(x + vh) \right).$$

By virtue of Duhamel's principle, the solution to system (1.0.22) is the limit, as $t \to \infty$, of the integral (in the sense of $\oplus$)

$$\hat{R}_h(t)\,\Phi_h(x) = \int_{\oplus}^{[0,t]} W_h(t,\ \tau,\ x)\,d\tau = \int_{\oplus}^{[0,t]} (\hat{L}_{h,V})^{t-\tau}\tilde{\Phi}_h(x)\,d\tau = \max_{0\leqslant\tau\leqslant t} W_h(t,\ \tau,\ x),$$

where $W_h(t,\ \tau,\ x)$ is the solution to the Cauchy problem on the segment $\tau \in [0,\ t]$, $\tau = lh$, for the homogeneous equation

$$W_h(t+h,\tau,x) = \hat{L}_{h,V} W_h(t,\tau,x),\ W_h(t,\tau,x)|_{t=\tau} = \tilde{\Phi}_h(x), \tag{1.0.24}$$

corresponding to (1.0.22). The function $f_h(t,\ x) = \hat{R}_h(t)\,\tilde{\Phi}_h(x)$ satisfies the equation

$$f_h(t+h,\ x) = \hat{L}_{h,V}\, f_h(t,\ x) \oplus \tilde{\Phi}_h(x),\ f_h(0,\ x) = \mathbb{O}. \tag{1.0.25}$$

Employing the obvious commutation relations for the A-valued σ_i-matrices, namely,

$$\sigma_1^2 = \sigma_4^2 = \hat{\mathbb{O}},\quad \sigma_3^2 = \sigma_3,\quad \sigma_2^2 = \sigma_2,$$

$$\sigma_i\sigma_j = \hat{\mathbb{O}},\quad (i,\ j) \in \{(1,\ 3),\ (2,\ 1),\ (2,\ 3),\ (3,\ 2),\ (3,\ 4),\ (4,\ 2)\},$$

$$\sigma_1\sigma_2 = \sigma_1,\ \sigma_4\sigma_3 = \sigma_4,\ \sigma_1\sigma_4 = \sigma_3,\ \sigma_2\sigma_4 = \sigma_4,\ \sigma_3\sigma_1 = \sigma_1,\ \sigma_4\sigma_1 = \sigma_2,$$

where $\hat{\mathbb{O}}$ is the "null" matrix, $\hat{\mathbb{O}} = [\delta_{ij}]$, $\delta_{ij} = \mathbb{O}$, and the operator methods developed in [1.13] and, in particular, the formulas of [1.14] and [1.15], both $(\hat{L}_{h,V})^{t-\tau}$ and $\hat{R}_h(t)$ can be calculated. The following equation holds true

$$\hat{R}_h(t)\,\psi(x) = htt_0 + \max_{0\leqslant l\leqslant k} \begin{pmatrix} \max\limits_{\substack{\zeta_1+\zeta_2=k-l-1\\ \zeta_i\in\mathbb{Z}_+}} \max\{\psi^1(x - h(\zeta_1+1)v_2 - h\zeta_2 v_3), \\ \psi^2(x - hv_4 - h\zeta_1 v_2 - h\zeta_2 v_3)\} \\ \max\limits_{\substack{\zeta_1+\zeta_2=k-l-1\\ \zeta_i\in\mathbb{Z}_+}} \max\{\psi^1(x - h\zeta_1 v_2 - h\zeta_2 v_3), \\ \psi^2(x - h\zeta_1 v_2 - h(\zeta_2+1)v_4)\} \end{pmatrix}, \tag{1.0.26}$$

where $t = kh,\ k - l \geqslant 0,\ k,\ l \in \mathbb{Z}_+$, and $\psi = (\psi^1,\ \psi^2)^{\mathrm{T}}$.

We denote by $\tilde{s}(x)$ the generalized solution to the stabilization Cauchy problem for the limiting generalized Hamilton-Jacobi equation to which (1.0.25) converges on smooth functions:

$$\tilde{s}(x) = s - \lim_{t\to\infty} u(x, t), \quad u(x, t) = (u^1, u^2)^{\mathrm{T}},$$

here $u(x, t)$ is the solution to the equation

$$\frac{\partial}{\partial t}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \max\{t_0 - \partial u^1/\partial x, \ t_0 - \partial u^2/\partial x - \partial u^2/\partial y\} \\ \max\{t_0 - u^1, \ t_0 - \partial u^2/\partial y\} \end{pmatrix}$$

with the initial data $u(x, t)\,|_{t=0} = \Phi_0(x) = (\Phi_0^1(x), \Phi_0^2(x))^{\mathrm{T}}$.

As in Example 3, formula (1.0.25) combined with $k \to \infty$, $kh \to t$, and $t \to \infty$ implies that for every smooth A-valued vector function $\varphi(x) = (\varphi^1(x), \varphi^2(x))$ we have

$$\begin{aligned}\langle \tilde{s}(x), \varphi(x)\rangle_{\oplus} &= \lim_{h\to 0}\lim_{t\to\infty} \langle \hat{R}_h(t)\, \tilde{\varphi}_h(x), \ \varphi(x)\rangle_{\oplus} \\ &= \lim_{h\to 0}\lim_{t\to\infty} \langle \tilde{\varphi}_h(x), \hat{R}_h^*(t)\, \varphi(x)\rangle = \langle \varphi_0(x), g_0(x)\rangle_{\oplus}, \quad (1.0.27)\end{aligned}$$

where

$$g_0(x) = \begin{Vmatrix} t_0 + \sup\limits_{\eta_i \in R_+^1} (t_0(\eta_1 + \eta_2) + \max\{\varphi^1(x + \eta_1 v_2 + \eta_2 v_3), \\ \qquad \varphi^2(x + \eta_1 v_2 + \eta_2 v_3)\}) \\ t_0 + \sup\limits_{\eta_i \in R_+^1} (t_0(\eta_1 + \eta_2) + \max\{\varphi^1(x + \eta_1 v_2 + \eta_2 v_3), \\ \qquad \varphi^2(x + \eta_1 v_2 + \eta_2 v_3)\}) \end{Vmatrix}.$$

Formula (1.0.27) defines, for any right-hand side $\Phi_0 = (\Phi_0^1, \Phi_0^2)^{\mathrm{T}}$, the generalized function that is the limit, as $h \to 0$ (in the above sense), of the initial discrete optimization problem (1.0.19).

1.1 The Theory of Linear Equations in Semi-modules

One of the main difficulties in solving optimization problems stems from the fact that these problems usually lead to nonlinear differential equations. However, it has been established that if we take function spaces with values in certain semi-rings and introduce definite nonarithmetic operations, many nonlinear equations important for applications become linear. This fact suggests the need for a thorough investigation of the corresponding axioms. This is done in the present chapter and is illustrated by numerous examples.

The approach developed here enables setting up and evaluating integrals of a special type. We prove the theorem that the integral of a function semi-measurable below is independent of the choice of the continuation of the idempotent measure. For the generalized Hamilton-Jacobi and Bellman equations (i.e. generalized in the new sense) considered in this chapter, we find the analogs of Duhamel's principle and the Fredholm alternative theorems.

1.1.1 Introduction

In this section we employ the examples of nonlinear Burgers and Hamilton-Jacobi equations to introduce the nonarithmetic operations of addition and multiplication with respect to which the resolving operators of the equations prove to be linear.

Let us start with the heat equation

$$\frac{\partial u}{\partial t}=\frac{h}{2}\,\frac{\partial^2 u}{\partial x^2}, \quad x\in R^1,\ t>0, \tag{1.1.1.1}$$

where h is a positive parameter. The linear combination

$$u=\lambda_1 u_1+\lambda_2 u_2 \tag{1.1.1.2}$$

of solutions u_1 and u_2 to Eq. (1.1.1.1) is also a solution to this equation.

Next we introduce

$$u=\exp\{-w(x,\ t)/h\} \tag{1.1.1.3}$$

in Eq. (1.1.1.1) and arrive at a nonlinear equation:

$$\frac{\partial w}{\partial t}+\frac{1}{2}\left(\frac{\partial w}{\partial x}\right)^2-\frac{h}{2}\,\frac{\partial^2 w}{\partial x^2}=0, \tag{1.1.1.4}$$

which is known as Burgers's equation.[1] Obviously, solution u_i $(i=1,\ 2)$ to Eq. (1.1.1.1) corresponds to solution $w_i=-h\ln u_i$ $(i=1,\ 2)$ to Eq. (1.1.1.4). Then solution (1.1.1.2) to Eq. (1.1.1.1) corresponds to the following solution to Eq. (1.1.1.4):

$$w=-h\ln\left(\exp\{(-w_1+\mu_1)/h\}+\exp\{(-w_2+\mu_1)/h\}\right),$$

where $\mu_i=-h\ln\lambda_i$ $(i=1,\ 2)$.

This implies that Eq. (1.1.1.4) is also linear; however, it is linear in the function space where semi-group operations have been introduced, namely, the "sum"

$$a\oplus b=-h\ln\left(\exp\{-a/h\}+\exp\{-b/h\}\right)$$

and the "product"

$$a\odot\lambda=a+\lambda.$$

[1] Ordinarily, this name is given to the equation obtained from Eq. (1.1.1.4) by differentiating with respect to x and substituting v for $\partial w/\partial x$.

The substitution $w = -h \ln u$ maps zero into ∞ and identity into zero. Thus, the semi-group zero in this new space is ∞, or $\mathbb{O} = \infty$, and the semi-group identity is the common zero, or $\mathbb{1} = 0$. The function space with the operations $\oplus$ and $\odot$ and the added zero $\mathbb{O}$ and identity $\mathbb{1}$ is isomorphic to the common function space with common multiplication and addition.

We can therefore assume that we have become accustomed to the new operations $\oplus$ and $\odot$ and that Eq. (1.1.1.4) is linear.

This example is quite simple, and we need not learn the new arithmetic operations since by substitution of functions we can transform from Eq. (1.1.1.4) to Eq. (1.1.1.5), which is linear in the common sense. However, it may so happen that equations interpreted in the sense of the new operations of addition and multiplication cannot be reduced by a substitution to the common linear case. It then becomes expedient to consider shifting the methods developed for linear equations to Eq. (1.1.1.4).

In the space of functions with values in the ring

$$a \oplus b = -h\,(\exp\{-a/h\} + \exp\{-b/h\}), \quad \lambda \odot b = \lambda + b,$$

we introduce the scalar product as follows:

$$(w_1,\ w_2) = -h \ln \int \exp\{-(w_1 + w_2)/h\}\, dx,$$

which, as we will show, is bilinear in this space, that is,

$$(a \oplus b,\ c) = (a,\ c) \oplus (b,\ c), \quad (\lambda \odot a,\ c) = \lambda \odot (a,\ c).$$

Indeed,

$$\begin{aligned}
(a \oplus b,\ c) &= -h \ln \Big(\int \exp\{-(1/h)\,[-h \ln (\exp\{-a/h\} \\
&\quad + \exp\{-b/h\} + c)]\}\, dx \Big) \\
&= -h \ln \Big(\int (\exp\{-a/h\} + \exp\{-b/h\}) \exp\{-c/h\}\, dx \Big) \\
&= -h \ln \Big(\int \exp\{-(a+c)/h\}\, dx \\
&\quad + \int \exp\{-(b+c)/h\}\, dx \Big) = (a,\ c) \oplus (b,\ c),
\end{aligned}$$

$$\begin{aligned}
(\lambda \odot a,\ c) &= -h \ln \Big(\int \exp\{-(a+\lambda)/h\} \exp\{-c/h\}\, dx \Big) \\
&= -h \ln \Big(\exp\{-\lambda/h\} \int \exp\{-(a+c)/h\}\, dx \Big) \\
&= \lambda - h \ln \int \exp\{-(a+c)/h\}\, dx = \lambda \odot (a,\ c).
\end{aligned}$$

An example of a hermitian (or self-adjoint) operator in this space is the operator

$$L\colon\ w \to w \odot (-h \ln ((w')^2/h^2 - (w'')/h)).$$

Let us check the hermiticity of this operator. We have

$$\begin{aligned}(w_1,\ Lw_2) &= -h \ln \int \exp\{-(w_1 + Lw_2)/h\}\, dx \\ &= -h \ln \int \exp\{-(1/h)(w_1 + w_2 - h \ln ((w_2')^2/h^2 - w_2''/h))\}\, dx \\ &= -h \ln \int \exp\{-w_1/h\} \exp\{-w_2/h\} [(w_2')^2/h^2 - w_2''/h]\, dx \\ &= -h \ln \int \exp\{-w_1/h\} \frac{\partial^2}{\partial x^2} \exp\{-w_2/h\}\, dx \\ &= -h \ln \int \left(\frac{\partial^2}{\partial x^2} \exp\{-w_1/h\}\right) \exp\{-w_1/h\}\, dx \\ &= -h \ln \int \exp\{-w_1/h\} [(w_1')^2/h^2 - w_1''/h] \exp\{-w_2/h\}\, dx \\ &= (Lw_1,\ w_2).\end{aligned}$$

We can also easily verify the linearity of operator L in the sense of operations $\oplus$ and $\odot$.

Let us construct the resolving operator of the Burgers equation, $\mathcal{L}_t\colon w_0 \to w$, where w is the solution to Eqs. (1.1.1.4) satisfying the initial data $w|_{t=0} = w_0$. The solution to Eq. (1.1.1.1.) satisfying the initial condition $u|_{t=0} = u_0$ has the form

$$u(x,\ t) = \frac{1}{\sqrt{2\pi h t}} \int \exp\{-(x-\xi)^2/(2th)\}\, u_0(\xi)\, d\xi.$$

Allowing for the fact that $u = \exp\{-w/h\}$, $w = -h \ln u$, we obtain the resolving operator $\mathcal{L}_t$ for the Burgers equation:

$$\mathcal{L}_t w_0 = -h \ln \left(\frac{1}{\sqrt{2\pi h t}} \int \exp\{-(x-\xi)^2/2th + w_0(\xi)/h\}\, d\xi\right).$$

Let us show that formally $\mathcal{L}_t$ is hermitian with respect to the new scalar product. Indeed,

$$\begin{aligned}(w_1,\ \mathcal{L}_t w_2) &= -h \ln \left(\frac{1}{\sqrt{2\pi h t}} \int \exp\left\{-\frac{1}{h}\left(w_1 - h \ln \left(\frac{1}{\sqrt{2\pi h t}}\right.\right.\right.\right. \\ &\quad \times \left.\left.\left.\left.\int \left(d\xi \exp\left\{-\frac{(x-\xi)^2}{2th} + \frac{w_2(\xi)}{h}\right)\right\}\right)\right)\right\} dx\right) \\ &= -h \ln \left(\frac{1}{\sqrt{2\pi h t}} \int \exp\left\{-\frac{1}{h}\left(w_1(x)\right.\right.\right. \\ &\quad \left.\left.\left. + h \ln \left(\exp\left\{ht \frac{\partial^2}{\partial x^2}\right\} \exp\{-w_2(x)/h\}\right)\right)\right\} dx\right)\end{aligned}$$

$$= -h \ln \Big(\frac{1}{\sqrt{2\pi ht}} \int \exp\{-w_1(x)/h\} \Big(\exp\Big\{ht \frac{\partial^2}{\partial x^2}\Big\}$$

$$\times \exp\{-w_2(x)/h\}\Big) dx\Big)$$

$$= -h \ln \Big(\frac{1}{\sqrt{2\pi ht}} \int \exp\{-w_2(x)/h\} \Big(\exp\Big\{ht \frac{\partial^2}{\partial x^2}\Big\}$$

$$\times \exp\{-w_1(x)/h\}\Big) dx\Big)$$

$$= -h \ln \Big(\frac{1}{\sqrt{2\pi ht}} \int \exp\Big\{-\frac{1}{h}\Big(w_2(x)$$

$$-h \ln\Big(\frac{1}{\sqrt{2\pi ht}} \int d\xi \exp\Big\{-\Big(\frac{(x-\xi)^2}{2th}$$

$$+ \frac{w_1(\xi)}{h}\Big)\Big\}\Big)\Big)\Big\} dx\Big) = (w_2, \mathscr{L}_t w_1).$$

If we send h to zero, the Burgers equation $2w_t + (w_x)^2 - hw_{xx} = 0$ transforms into the Hamilton-Jacobi equation

$$\frac{\partial S}{\partial t} + \frac{1}{2}\left(\frac{\partial S}{\partial x}\right)^2 = 0. \tag{1.1.1.5}$$

The operation of "addition" $a \oplus b = -h\,(\ln(\exp\{-a/h\} + \exp\{-b/h\}))$ transforms, as $h \downarrow 0$, into $a \oplus b = \min(a, b)$. The "product" (or the operation of "multiplication") does not depend on h. Hence, we again have $a \odot \lambda = a + \lambda$. The two operations are distributive, that is, $(a \oplus b) \odot c = (a \odot c) + (b \odot c)$.

Let us demonstrate that the Hamilton-Jacobi equation is linear in the function space with the operations $a \oplus b = \min(a, b)$ and $a \odot \lambda = a + \lambda$. Let us consider the Cauchy problem for the Hamilton-Jacobi equation:

$$\frac{\partial S}{\partial t}(x, t) + H\left(\frac{\partial S}{\partial x}, x, t\right) = 0, \tag{1.1.1.6}$$

$$S(x, 0) = S_0(x), \quad 0 \leqslant t \leqslant T, \quad x \in R^n,$$

where $H(p, x, t)$ is a function (the Hamiltonian that is smooth on $R^{2n} \times [0, T]$ and has a positive definite second derivative with respect to p. Suppose that $\mathscr{L}(v, x, t)$ is the Lagrangian that corresponds to the Hamiltonian $H(p, x, t)$, or

$$\mathscr{L}(v, x, t) = \max_{p \in R^n} (vp - H(p, x, t)).$$

If for every $t \in [0, T]$ the Lagrange manifold Λ^t obtained from the graph Λ^0 of the differential of the function S_0 through the action of the phase flux g^t_H generated by the system of Hamilton's equa-

tions

$$\frac{\partial p}{\partial t} = -\frac{\partial H}{\partial x}(p, x, t), \quad \frac{dx}{dt} = \frac{\partial H}{\partial p}(p, x, t)$$

is mapped diffeomorphically on the subspace $p \equiv 0$, then there exists a smooth and unique solution to problem (1.1.1.6). Here $S(x, t)$ is the minimum of the functional

$$J(y(\cdot)) = S_0(y(0)) + \int_0^t \mathscr{L}(\dot{y}(\tau), y(\tau), \tau)\, d\tau \qquad (1.1.1.7)$$

on the set $\Phi_{x,t}$ of piecewise smooth functions $y(\cdot)$ defined on $[0, t]$ and satisfying the condition $y(t) = x$. This justifies the following definition: a function S whose value at a point (x, t) is the greatest lower bound of the functional J on the set $\Phi_{x,t}$ is called the generalized solution to the Cauchy problem (1.1.1.6).

With such a definition problem (1.1.1.6) has a generalized solution for every function S_0 with values on the extended real axis $\overline{R^1}$. Thus, we have defined the resolving operator $\mathscr{A}_H$: $S_0 \to S$ for problem (1.1.1.6).

Suppose that $\oplus$ is the operation of the pointwise minimum of two functions with values in R^1: $S_1 \oplus S_2(x) = \min(S_1(x), S_2(x))$.

Theorem 1.1.1.1 *Operator $\mathscr{A}_H$ is additive with respect to the operation $\oplus$.*

Proof. Suppose that

$$\sigma_\xi(x) = \begin{cases} 0 & \text{if } x = \xi \\ \infty & \text{if } x \neq \xi. \end{cases}$$

Then $\mathscr{A}_H \delta_\xi(x, t)$ is the minimal value of the functional $\int_0^t \mathscr{L}(\dot{y}(\tau), y(\tau), \tau)\, d\tau$ in the class $\Phi_{\xi,x,t}$ of piecewise smooth functions $y(\cdot)$ that satisfy the boundary conditions $y(0) = \xi$ and $y(t) = x$. Minimizing functional (1.1.1.7) in the functions $y(\cdot) \in \Phi_{x,t}$ that satisfy the initial condition $y(0) = \xi$ and then calculating the greatest lower bound in $\xi \in R^n$, we arrive at the formula

$$\mathscr{A}_H S_0(x, t) = \inf_{\xi \in R^n} (\mathscr{A}_H \delta_\xi(x, t) + S_0(\xi)).$$

Hence,

$$\begin{aligned}(\mathscr{A}_H S_{01} \oplus \mathscr{A}_H S_{02})(x, t) &= \min(\inf_\xi (\mathscr{A}_H \delta_\xi(x, t) \\ &\quad + S_{01}(\xi)), \inf_\xi (\mathscr{A}_H \delta_\xi(x, t) + S_{02}(\xi))) \\ &= \inf_\xi \min(\mathscr{A}_H \delta_\xi(x, t) + S_{01}(\xi), \mathscr{A}_H \delta_\xi(x, t) \\ &\quad + S_{02}(\xi)) \\ &= \inf_\xi (\mathscr{A}_H \delta_\xi(x, t) + \min(S_{01}(\xi), S_{02}(\xi))) \\ &= \mathscr{A}_H(S_{01} \oplus S_{02})(x, t).\end{aligned}$$

A similar theorem holds true for the Bellman equation

$$\max_{v\in V(x,t)}\left(\frac{\partial S}{\partial t}(x,\ t)+v\frac{\partial S}{\partial x}(x,\ t)-\mathcal{L}\ (v,\ x,\ t)\right)=0$$

with the initial condition $S\ (x,\ 0) = S_0\ (x)$, provided that the generalized solution of this Cauchy problem is understood in the following sense:

$$S\ (x,\ t)=\inf_{\substack{y(t)=x\\ \dot{y}\ (t)\in V(x,t)}}\left(S_0\ (y\ (0))+\int_0^t \mathcal{L}\ (\dot{y}\ (\tau),\ y\ (\tau),\ \tau)\ d\tau\right).$$

1.1.2 The Metric and Structure on a Semi-ring

In this section we consider the axiomatics of abstract semi-group operations of addition and multiplication in a partially ordered metric semi-ring. Operations performed on points of a semiring generate in a natural manner operations performed on functions with values in the semi-ring A, and this makes it possible to define an A-valued scalar product and an A-valued delta function. We also discuss some examples.

Suppose that A is an Abelian semi-ring defined by the commutative semi-group operations of "addition" $\oplus$ and "multiplication" with neutral elements $\mathbb{0}$ and $\mathbb{1}$, respectively,

$$\mathbb{0}\odot a=\mathbb{0},\qquad \mathbb{1}\odot a=a,\quad \mathbb{0}\oplus a=a,$$

and suppose that these operations obey the distributivity condition on A, that is, $a\odot(b\oplus c)=(a\odot b)\oplus(a\odot c)$.

We will assume that the partial ordering relation $\leqslant$ has been defined on A and that this relation has the following properties:

$$a\odot a\geqslant \mathbb{0}\quad \forall\, a\in A;$$
$$\{a\leqslant b\}\Rightarrow\{a\oplus c\leqslant b\oplus c\quad \forall\, c\in A\};$$
$$\{a\geqslant b\}\Rightarrow\{a\odot c\geqslant b\odot c\quad \forall\, c\geqslant \mathbb{0}\}.$$

If elements a and b are not congruent, we will write $a\varnothing b$.

Example 1. Let us take the operation max for the operation of semi-group addition $\oplus$ on the extended real axis $A=R\cup\{-\infty\}$ and the operation of common addition for the operation of semi-group multiplication $\odot$. Zero is the semi-group's identity $\mathbb{1}$ in this semiring, while the adjoined point $-\infty$ is the semi-group zero $\mathbb{0}$. Indeed,

$$\mathbb{0}\odot a=a-\infty=-\infty=\mathbb{0},$$
$$\mathbb{1}\odot a=a+0=a,\ \mathbb{0}\oplus a=\max\{a,\ -\infty\}=a,$$
$$a\cdot(b\oplus c)=a+\max\ (b,\ c)=\max\ (a+b,\ a+c)$$
$$=(a\odot b)\oplus(a\odot c).$$

The ordering relation on A coincides with the natural ordering relation on the real axis, with $a \geqslant \mathbb{0}$ for every $a \in A$.

Example 2. Let us take the operation min for the operation of semi-group addition $\oplus$ on the extended real axis $A = R^1 \cup \{\infty\}$ and the operation of common addition for the operation of semi-group multiplication $\odot$. Point 0 is the semi-group identity $\mathbb{1}$, while point $\{+\infty\}$ is the semi-group zero $\mathbb{0}$:

$$\mathbb{0} \odot a = a + \infty = \infty = \mathbb{0}, \qquad \mathbb{1} \odot a = a + 0 = a,$$
$$\mathbb{0} \oplus a = \min\{\infty, a\} = a,$$
$$a \odot (b \oplus c) = a + \min(b, c) = \min(a + b, a + c)$$
$$= (a \odot b) \oplus (a \odot c).$$

The ordering relation on A is opposite to the natural ordering relation on the real axis. At the same time here $a \geqslant \mathbb{0}$ for every $a \in A$, as in the previous example.

Example 3. Let us take the semi-ring $A = R_+ \cup \{\infty\}$ generated by the operations $\oplus = \max$ and $\odot = \min$ the neutral elements $\mathbb{0} = 0$ and $\mathbb{1} = \infty$:

$$\mathbb{0} \odot a = \min\{0, a\} = \mathbb{0}, \quad \mathbb{1} \odot a = \min\{a, \infty\} = a,$$
$$\mathbb{0} \oplus a = \max\{0, a\} = a,$$

$$\begin{aligned} a \odot (b \oplus c) &= \min\{a, \max\{b, c\}\} \\ &= \min\{a, \max\{\min\{a, b\}, \min\{a, c\}\}\} \\ &= \max\{\min\{a, b\}, \min\{a, c\}\} \\ &= (a \odot b) \oplus (a \odot c). \end{aligned}$$

The ordering relation on A coincides with the natural one and $a \geqslant \mathbb{0}$ for every $a \in A$.

Example 4. Suppose that $A = R_+ \cup \{\infty\}$, $\oplus = \min$, $\odot = \max$, $\mathbb{0} = \infty$, and $\mathbb{1} = 0$. Then

$$\mathbb{0} \odot a = \max\{\infty, a\} = \mathbb{0}, \qquad \mathbb{1} \odot a = \max\{0, a\},$$
$$\mathbb{0} \oplus a = \min\{\infty, a\} = a,$$

$$\begin{aligned} a \odot (b \oplus c) &= \max\{a, \min\{b, c\}\} \\ &= \max\{a, \min\{\max\{a, b\}, \max\{b, c\}\}\} \\ &= \min\{\max\{a, b\}, \max\{b, c\}\} \\ &= (a \odot b) \oplus (a \odot c), \end{aligned}$$

and $a \geqslant \mathbb{0} \quad \forall a \in A$.

Example 5. Suppose that $A = R_+$, $\oplus = \max$, $\mathbb{0} = 0$, $\odot = \times$, and $\mathbb{1} = 1$. Then

$$\mathbb{0} \odot a = 0 \times a = \mathbb{0}, \qquad \mathbb{1} \odot a = 1 \times a = a,$$

$$\mathbb{0} \oplus a = \max \{0, a\} = a,$$

$$a \odot (b \oplus c) = a \times \max \{b, c\} = \max \{a \times b, a \times c\}$$
$$= (a \odot b) \oplus (a \odot c).$$

Let us now assume that topology on A is given via the metric $\rho: A \times A \to [0, \infty)$ that satisfies the standard axioms:

$$(1)\ \rho(a, a) = 0, \quad (2)\ \rho(a, b) = \rho(b, a)$$
$$(3)\ \rho(a, b) + \rho(b, c) \geqslant \rho(a, c)$$
$$(4)\ \rho(a, b) = 0 \Leftrightarrow a = b.$$

Set A is said to be a complete structure if any subset $a \in A$ bounded above (below) contains the least upper (greatest lower) bound. In what follows we assume that the semi-ring A is complete both as a structure and as a metric space. In addition, we assume that metric and structure are concordant in the sense of convergence, that is, $\rho(a_n, a) \to 0$ implies $\lim\limits_{N\to\infty} \sup\limits_{n \geqslant N} a_n = \lim\limits_{N\to\infty} \inf\limits_{n \geqslant N} a_n = a$, and vice versa.

The following metrics are examples of metrics on $R \cup \{-\infty\}$:

$$\rho^+_{\exp}(a, b) = \exp\{\max(a, b)\} - \exp\{\min(a, b)\},$$

$$\rho_{\tan^{-1}}(a, b) = \tan^{-1}\{\max(a, b)\} - \tan^{-1}\{\min(a, b)\}.$$

For the metric on $(R \cup \{-\infty\})^n$ we can take

$$\rho^+_{\exp}(a, b) = \sum_{j=1}^{n} \rho^+_{\exp}(a_j, b_j),$$

$$\rho_{\tan^{-1}}(a, b) = \sum_{j=1}^{n} \rho_{\tan^{-1}}(a_j, b_j),$$

where $a = (a_1, \ldots, a_n)$, and $b = (b_1, \ldots, b_n)$.

An example of a partial ordering relation in $(R \cup \{-\infty\})^h$ is the relation $(a_1, \ldots, a_n) \geqslant (b_1, \ldots, b_n)$ if $a_i \geqslant b_i\ \forall i$. The set $(R \cup \{-\infty\})^h$ equipped with such an ordering relation is a complete structure concordant with the metrics $\rho_{\exp}$ and $\rho_{\tan^{-1}}$, with $a \oplus b = \max(a, b)$.

The natural metrics for the semi-ring $A = (R \cup \{\infty\})^n$ are $\rho^-_{\exp}$ and $\rho_{\tan^{-1}}$:

$$\rho^-_{\exp}(a,\ b) = \sum_{j=1}^{n} \rho^-_{\exp}(a_j,\ b_j),$$

$$\rho^-_{\exp}(a_j,\ b_j) = \exp(-\min\{a_j,\ b_j\}) - \exp(-\max\{a_j,\ b_j\}).$$

The metrics $\rho^{\pm}_{\exp}$ and $\rho_{\tan^{-1}}$ ensure that the corresponding metric spaces are precompact, that is, that there exists a finite ε-mesh on any bounded subset. Indeed, on the set $M = \{x\colon \rho^+_{\exp}(x,\ \mathbb{0}) \leqslant a\}$ there exists a finite ε-mesh with respect to the metric $\rho^+_{\exp}$: $N = [a/\varepsilon] + 1$, $d^\varepsilon_N = \mathbb{0} = -\infty$, $d^\varepsilon_K = \ln(a - K\varepsilon)$, $0 \leqslant K < N$. On set $R \cup \{\pm\infty\}$, the ε-mesh with respect to metric $\rho_{\tan^{-1}}$ is the set of points $\{g^\varepsilon_k\}^N_1$, with $N = [\pi/\varepsilon] + 1$, $g^\varepsilon_0 = \mathbb{0} = -\infty$, $g^\varepsilon_k = \tan(-\pi/2 + k\varepsilon)$, $0 \leqslant k < N$, $g^\varepsilon_N = \mathbb{1} = \infty$.

Clearly, Axioms (1), (2), and (4) are satisfied for the above-mentioned metrics. Let us now verify the triangle property (3) for metric $\rho_{\exp}$ and $\rho_{\tan^{-1}}$. It is sufficient to consider only the one-dimensional case, since the verification of (3) for a multidimensional case can be reduced to one-dimensional verification. In one dimension we have

$$\begin{aligned} \rho(a,\ b) + \rho(b,\ c) &= \rho(a,\ c) \quad \text{if} \quad b \in [a,\ c], \\ \rho(b,\ c) &\geqslant \rho(a,\ c) \quad \text{if} \quad b \leqslant a, \\ \rho(a,\ b) &\geqslant \rho(a,\ c) \quad \text{if} \quad b \geqslant c, \end{aligned}$$

whereby (3) is sure to be true in all cases.

For sets that are simultaneously structures and topological spaces we can define the ordering relation $>$ as follows: $a > b$ if $a \in \operatorname{int}\{c\colon c \geqslant b\}$. Clearly, here the ordering relation $>$ obeys the transitivity law: if $a > b$ and $b > c$, then $a > c$. In the discussions below we will assume that $\mathbb{1} > \mathbb{0}$ and that for every a and c such that $a > c$ there exists a b: $a > b > c$.

We start by giving a formal definition of a space with values in a semi-ring without employing the concept of idempotency. In Section 1.1.3 we introduce the concept of an idempotent measure, which will enable us to give a complete description of the functions belonging to this space.

We denote by C_0 the set of continuous A-valued functions on a locally compact space X that are nonzero only inside certain compact metric spaces $K \subset X$.

On set C_0 we consider the structure A, which is a semi-module with respect to the operations $(\varphi \oplus \psi)(x) = \varphi(x) \oplus \psi(x)$ and $(a \odot \varphi) \times (x) = a \odot \varphi(x)$ with a zero $\varphi(x) = \mathbb{0}$, an identity $\varphi(x) = \mathbb{1}$, and an ordering relation $\varphi \leqslant \psi \Rightarrow \varphi(x) \leqslant \psi(x) \quad \forall x \in X$. We equip the ordered semi-module $C_0(X)$ with a uniform structure de-

fined by the deviation

$$\rho(\varphi, \psi) = \sup_{x \in X} \rho(\varphi(x), \psi(x))$$

with respect to which the semi-module is topological and possesses uniform convergence $\varphi_n \to \varphi$ if $\rho(\varphi_n, \varphi) \downarrow 0$. We will say that a function $\psi(x)$ is semi-continuous below if the condition $a < \psi(x_0)$ at point x_0 implies the existence of a (compact) neighborhood $V \subset X$ of this point in which (neighborhood) $a < \psi(x)$ $\forall x \in V$. For the product $\varphi \odot \psi$ of functions φ and ψ we take the pointwise product $(\varphi \odot \psi)(x) = \varphi(x) \odot \psi(x)$. Here are some definitions of the scalar product.

Example 6. Let us consider the space with values in a semi-ring generated by the operations (inf, +). The scalar product in such a space has the form $(\varphi, \psi) = \inf_x (\varphi(x) + \psi(x))$.

An example of a sequence converging to the delta function is $\delta_n(x - \xi) = n(x - \xi)^2$. Indeed, $\lim_{n\to\infty} \inf_x (n(x - \xi)^2 + \varphi(x)) = \varphi(\xi)$.

Example 7. Let us consider the space with values in a semi-ring generated by the operations (sup, +). The scalar product in such a space has the form $(\varphi, \psi) = \sup_x (\varphi(x) + \psi(x))$.

An example of a sequence converging to the delta function is $\delta_n(x - \xi) = -n(x - \xi)^2$. Indeed, $\lim_{n\to\infty} \sup_x (-n(x - \xi)^2 + \varphi(x)) = \varphi(\xi)$.

Example 8. Let us consider the semi-ring $A = R_+$ generated by the operations (min, max). In the space of functions with values in this semi-ring the scalar product has the form $(\varphi, \psi) = \min_x \max(\varphi(x), \psi(x))$.

An example of a sequence converging to the delta function is $\delta_n(x - \xi) = n(x - \xi)^2$. Indeed, $\lim_{n\to\infty} \inf_x ((x - \xi)^2 n, \varphi(x)) = \varphi(\xi)$. Since $\varphi(x) \geqslant 0$, at $x = \xi$ we have $\max(n(x - \xi)^2, \varphi(x)) = \varphi(\xi)$.

Example 9. Let us consider the space of functions with values in the semi-ring $A = R_+$ generated by the operations (max, min). The scalar product has the form $(\varphi, \psi) = \max_x \min(\varphi(x), \psi(x))$.

An example of a sequence converging to the delta function is $\delta_n = 1/[n(x - \xi)^2]$. Indeed,

$$\lim_{n\to\infty} \max_x \{\min(1/[n(x - \xi)^2], \varphi(x))\} = \varphi(\xi).$$

Since $\varphi(x) \geqslant 0$, at $x = \xi$ we have

$\min(1/[n(x - \xi)^2], \varphi(x)) = \min(0, \varphi(\xi)) = \varphi(\xi)$.

Example 10. In the space with values in the semi-ring generated by the operation (max, $\times$) the scalar product has the form $(\varphi, \psi) = \max \{\varphi(x), \psi(x)\}$.

An example of a sequence converging to the delta function is $\delta_n(x - \xi) = \exp\{-n(x - \xi)^2\}$. Indeed,

$$\lim_{n\to\infty} \max_x \{\exp\{-n(x-\xi)^2\},\ \varphi(x)\} = \varphi(\xi),$$

since

$$\lim_{n\to\infty} \exp\{-n(x-\xi)^2\} = \begin{cases} 0 & \text{if } x \neq \xi, \\ 1 & \text{if } x = \xi. \end{cases}$$

We will now consider the endomorphisms G of the semi-module $C_0(X)$, that is, the mappings G of $C_0(X)$ into itself that satisfy the condition $G(a \odot \varphi \oplus b \odot \psi) = a \odot G(\varphi) \oplus b \odot G(\psi)$ for all $\varphi, \psi \in C_0(X)$, with $a, b \in A$, and that are continuous with respect to uniform convergence, $\varphi_n \to \varphi \Rightarrow G(\varphi_n) \to G(\varphi)$. In what follows we call these endomorphisms operators.

The conjugate of the operator G with respect to the scalar product defined above, $(\cdot, \cdot)$, is the operator G^* that satisfies the condition $(G\varphi, \psi) = (\varphi, G^*\psi)$, with $\varphi, \psi \in C_0(X)$. The conjugate operator has values that are functions (generalized, generally speaking) and can be defined on any generalized function as a conjugate endomorphism of A, which is a semi-module of the generalized functions, with the endomorphism continuous in the topology of weak convergence of these functions.

1.1.3 Semi-group Additivity of Idempotent Measures

Here we will show how one can construct an integral if the addition operation is idempotent while the measure is not continuous at zero. The central role in such a construction is played by the property of semi-continuity of the functions being integrated.

The set functions considered in this section are similar to measures, that is, are of fixed sign and additive. To give a simple but interesting example, we equip the set $A = R \cup \{-\infty\}$ with the natural ordering relation $\geqslant$ and the semi-group operations $a \oplus b = \max\{a, b\}$ and $a \cdot b = a + b$. For the neutral elements of addition $\oplus$ and multiplication $\odot$ we take the elements $\mathbb{0} = -\infty$ and $\mathbb{1} = 0$, respectively. Then all real values prove to be nonnegative and the functions bounded above prove to be bounded and nonnegative in the sense of the structure of semi-ring A.

The set function

$$m(B) = \sup_{x\in B} f(x),\ f \in C(R^n),\ f \leqslant M < \infty, \tag{1.1.3.1}$$

with $B \in \mathscr{B}(R^n)$ a Borel σ-algebra on R^n, is defined for every function $f\colon R^n \to R$ and is a monotonic nonnegative countably

additive set function:

$$m\left(\bigcup_{i=1}^{\infty} B_i\right) = \sup_{x\in\bigcup B_i} f(x) = \sup_i \left(\sup_{x\in B_i} f(x)\right) = \bigoplus_1^{\infty} m(B_i). \tag{1.1.3.2}$$

If $f(x)$ is nonpositive in the ordinary sense, then $m: \mathscr{B}(R^n) \to (\mathbb{O}, \mathbb{1}]$. The idempotency of measure m lies in the fact that $m(B) = m(B) \oplus m(B)$.

Note that the σ-additivity in (1.1.3.2) occurs for intersecting sets B_i, too.

In contrast to classical (probability) σ-additive measures, the set function (1.1.3.1) is not continuous on empty sets. For example, suppose that $\{B_k\}_1^\infty$ is a sequence of open balls of radii $1/k$ with centers at the points $r_k = \{1/k, 0, 0, \ldots, 0\}$. Then, obviously, $\bigcap_1^\infty B_k = \varnothing$, but $\lim_{k\to\infty} m(B_k) = \lim_k \sup_{x\in B_k} f(x) = f(0) \neq 0$ for every continuous function $f(x)$. This result contradicts the well-known definition of continuity of measure on empty sets. More than that, it is clear that $m(\varnothing)$ depends on the choice of f and the sequence $\{B_k\}_1^\infty$ converging to an empty set, whereby $m(\varnothing)$ cannot have a definite value in principle. Roughly, different empty sets exist. This is closely related to the specific restrictions on the system of sets S on which the idempotent measure is given. The system S must be closed with respect to operation $\bigcup$, the union of sets. This operation generates the addition $\oplus$ of measures and does not lead to the emergence of empty sets; it can be used to continue an idempotent measure onto the class of sets that is closed with respect to (countable) unions.

In spite of the fact that idempotent measures cannot be subtracted (so that operations $\bigcap$ and $\setminus$ cannot be applied to the elements of S), the supply of sets in S should be sufficiently large. In particular, it would be desirable to have a system S that incorporates all the sets (except $\varnothing$) of a σ-algebra $\mathfrak{A}$ generated by an algebra Σ.

For this reason we will assume that either an idempotent measure is given on all nonempty sets of the σ-algebra $\mathfrak{A}$ or it can be defined on $\mathfrak{A}$ in such a manner that the conditions of σ-additivity, nonnegativity, and idempotency are met.

An important feature of an idempotent measure that puts such a measure apart from the common one is the existence of various σ-additive idempotent continuations from algebra Σ to the minimal σ-algebra $\mathfrak{A}$ generated by Σ. Suppose that Σ_R and Σ_D are two algebras of subsets of R generated by segments with rational and real end points, respectively. Both contain no isolated points and generate the same σ-algebra $\mathscr{B}(R)$. Since π is irrational, the four functions $f_i(x)$ (see Figure 1) equal to zero outside the segment with end points 3 and 4 have corresponding to them the same idempotent

measure on algebra Σ_R:

$$m_i^R(B) = \sup_{x \in B} f_i(x), \ B \in \Sigma_R, \ 1 \leqslant i \leqslant 4.$$

At the same time, on the algebra Σ_D the functions $m_i^{\mathscr{Z}}(B) = \sup_{x \in B} f_i(x)$, $B \in \Sigma_D$, are distinct. In particular, $m_1^{\mathscr{Z}}[\pi, 4] = a$, $m_2^{\mathscr{Z}}[\pi, 4] = m_4^{\mathscr{Z}}[\pi, 4] = b$, and $m_3^{\mathscr{Z}}[\pi, 4] = c$.

On algebra $\mathscr{B}(R)$, all measures $m_i(\beta) = \sup_{x \in \beta} f_i(x)$, $\beta \in \mathscr{B}(R)$, are distinct. For example, $m_1(\{\pi\}) = a$, $m_2(\{\pi\}) = b$, $m_3(\{\pi\}) = c$, and $m_4(\{\pi\}) = d$.

The restriction of measures m_i to algebra Σ_R coincides with m^R, and therefore each measure m_i is a continuation of measure m^R on the σ-algebra $\mathscr{B}(R)$. The arbitrariness of the choice of c and d indicates an infinitude of various continuations.

Suppose that we have chosen a continuation. Our immediate goal is to interpret $\sup(f(x) + \varphi(x))$ as a Lebesgue integral and to formulate the conditions under which the value of the integral does not depend on the choice of continuation. Suppose f and φ are two real functions on R, with φ measureable and bounded above. By $\{\varphi_i^\varepsilon\}_1^\infty$ we denote the ε-mesh on the set of values of φ, and by Q_i^ε and q_i^ε the following Borel sets:

$$Q_i^\varepsilon = \{x\colon \varphi(x) \geqslant \varphi_i^\varepsilon\},$$

$$q_i^\varepsilon = \{x\colon \varphi_i^\varepsilon \leqslant \varphi(x) < \varphi_i^\varepsilon + \varepsilon\},$$

$Q_i^\varepsilon \supset q_i^\varepsilon$. The sets Q_i^ε and q_i^ε cover the entire set of values of map φ, whereby

$$I = \sup_x (f(x) + \varphi(x)) = \sup_i (\sup_{Q_i^\varepsilon} (f + \varphi)) = \sup_i (\sup_{q_i^\varepsilon} (f + \varphi)).$$

Allowing for the fact that for the function φ the two-sided bound $\varphi_i^\varepsilon \leqslant \varphi(x) < \varphi_i^\varepsilon + \varepsilon$ is valid on q_i^ε, we get

$$\sup_i (\varphi_i^\varepsilon + \sup_{q_i^\varepsilon} f) \leqslant \sup_i (\varphi_i^\varepsilon + \sup_{Q_i^\varepsilon} f) \leqslant I$$

$$\leqslant \sup_i (\varphi_i^\varepsilon + \varepsilon + \sup_{q_i^\varepsilon} f) \leqslant \sup_i (\varphi_i^\varepsilon + \varepsilon + \sup_{Q_i^\varepsilon} f).$$

Thus,

$$\sup_i (\varphi_i^\varepsilon + \sup_{q_i^\varepsilon} f) \leqslant I \leqslant \sup_i (\varphi_i^\varepsilon + \sup_{q_i^\varepsilon} f) + \varepsilon,$$

$$\sup_i (\varphi_i^\varepsilon + \sup_{Q_i^\varepsilon} f) \leqslant I \leqslant \sup_i (\sup_{Q_i^\varepsilon} f + \varphi_i^\varepsilon) + \varepsilon.$$

This implies that

$$\sup_{x\in R^n}(f(x)+\varphi(x))=\lim_{\varepsilon\downarrow 0}\sup_i(\varphi_i^\varepsilon+\sup_{x\in q_i^\varepsilon} f)$$
$$=\lim_{\varepsilon\downarrow 0}\sup_i(\varphi_i^\varepsilon+\sup_{x\in Q_i^\varepsilon} f).$$

Employing the operations $\odot$ and $\oplus$ defined earlier, we can write the above result in the form of a sum:

$$I(\varphi)=\sup_{x\in R^n}(f(x)+\varphi(x))=\lim_{\varepsilon\downarrow 0}\bigoplus_{i=1}^{\infty}(\varphi_i^\varepsilon\odot m(Q_i^\varepsilon))$$
$$=\lim_{\varepsilon\downarrow 0}\bigoplus_{i=1}^{\infty}(\varphi_i^\varepsilon\odot m(q_i^\varepsilon))=\overset{\oplus}{\int_{R^n}}\varphi(x)\odot m(dx), \quad (1.1.3.3)$$

where $m(B)=\sup_{x\in B} f(x)$. The right-hand side of this equation can be called an idempotent integral, since it possesses the following property: $I(\varphi)\oplus I(\varphi)=I(\varphi)$.

For each continuation m_R to a σ-additive idempotent set function m we can define a procedure for the integration of measurable functions according to the scheme suggested above; in its main features this scheme coincides with the definition of a Lebesgue integral.

The integral of a simple bounded-above function $\varphi(x)$ that assumes values φ_i on the sets q_i is equal to

$$\overset{\oplus}{\int}\varphi(x)\odot m(dx)=\bigoplus_{1}^{\infty}(\varphi_i\odot m(q_i)). \quad (1.1.3.4)$$

Definition (1.1.3.4) allows the following important modification. Using the line of reasoning similar to the one employed in deriving Eq. (1.1.3.3), we can write

$$\bigoplus_{1}^{\infty}(\varphi_i\odot m(q_i))=\bigoplus_{1}^{\infty}(\varphi_i\odot m(Q_i)), \quad (1.1.3.5)$$

where $Q_i=\bigcup_{j\in J(i)} q_j$, $J(i)=\{j\colon \varphi_j\geqslant\varphi_i\}$. Indeed, $q_i\subseteq Q_i$, whereby $m(Q_i)\geqslant m(q_i)$ and the right-hand side of (1.1.3.5) is at least no smaller than the left-hand side.

On the other hand, employing the σ-additivity of the idempotent measure m, we can represent $m(Q_i)$ in the form of a sum:

$$m(Q_i)=\bigoplus_{j\in J(i)} m(q_j).$$

Hence,

$$\bigoplus_{1}^{\infty}(\varphi_i\odot m(Q_i))=\bigoplus_{i=1}^{\infty}(\varphi_i\odot(\bigoplus_{j\in J(i)} m(q_i))$$
$$\leqslant\bigoplus_{i=1}^{\infty}(\bigoplus_{j\in J(i)}(\varphi_j\odot m(q_j))).$$

The repetition of terms does not change the idempotent sum $\oplus$. This implies that the left-hand side of (1.1.3.5) is no smaller than the right-hand side. We have proved the validity of Eq. (1.1.3.5).

Let us define an idempotent Lebesgue integral for a measurable function φ bounded above and a function f (also bounded above) with which measure m is associated. For such an integral it is natural to take the limit of a sequence of idempotent integrals of a bounded sequence of simple functions converging to φ:

$$I(\varphi) = \lim_{\varepsilon \downarrow 0} \bigoplus_{1}^{\infty} (\varphi_i^\varepsilon \odot m(Q_i^\varepsilon)) = \int_{R^n}^{\oplus} \varphi(x) \odot m(dx).$$

The existence of the limit follows from the separability and local compactness of set R^n and also from the boundedness and monoto-

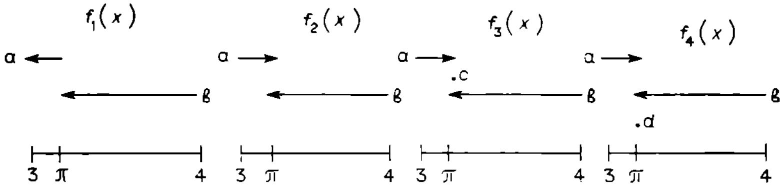


Fig. 1.1

nicity of the sequence of integrals of the simple functions for increasing sequences of ε-meshes. These properties of R^n and the sequence $I_\varepsilon(\varphi)$ will be used later in constructing a Lebesgue integral with values in a partially ordered space.

Let us now return to the question of the possibility of distinct continuation of an idempotent measure. Figure 1.1 shows that the measures $m_i(x)$ with the same restrictions to algebra Σ_R correspond to functions with a common upper bound $f_1(x)$ semi-continuous above

$$\begin{aligned} f_1(x) &= \inf \{\psi(x) \colon \psi(x) \geqslant f_i(x),\ \psi(x) \in C(R)\} \\ &\geqslant f_i(x), \quad i = 1, 2, 3, 4. \end{aligned}$$

Simple examples show that $\sup_x (f_i(x) + \varphi(x)) = \sup_x (f_1(x) + \varphi(x))$, $1 \leqslant i \leqslant 4$, for every function φ semi-continuous below.

Let us now consider two functions, φ_1 and φ_2. For $\varphi_1(x)$ we have $\sup_x (f_i(x) + \varphi_1(x)) = \max(\alpha + a, \beta + b)$, $i = 1, 2, 3, 4$, and $\varphi_1(x)$ is semi-continuous below. At the same time, for the function

$\varphi_2(x)$, which is semi-continuous above, we have

$$\sup (f_1 + \varphi_2) = \beta + a,$$
$$\sup (f_{2,4} + \varphi_2) = \max (\alpha + a, \beta + b),$$
$$\sup (f_3 + \varphi_2) = \max (\alpha + a, \beta + c),$$
$$\sup (f_5 + \varphi_2) = \max (\alpha + c, \beta + b),$$
$$\sup (f_5 + \varphi_2) = \beta + b.$$

The natural generalization of the above reasoning is summed up in the following

Theorem 1.3.1 *Suppose f and φ are two real functions on R^n, with φ semi-continuous below. Suppose also that the function*

$$(\mathrm{cl}\, f)(x) = \inf \{\psi(x) \colon \psi \geqslant f, \psi \in C(R^n)\} \tag{1.1.3.6}$$

is semi-continuous above. Then

$$\sup_x (f(x) + \varphi(x)) = \sup_x ((\mathrm{cl}\, f)(x) + \varphi(x)). \tag{1.1.3.7}$$

Operation cl *is known as the closure.*

Proof. We note first that $\mathrm{cl}\, f \geqslant f$ and that for every $\varepsilon > 0$ and an arbitrarily small open neighborhood U_x of a point x there is a point y such that $(\mathrm{cl}\, f)(x) - f(y) \leqslant \varepsilon$. True, since otherwise $(\mathrm{cl}\, f)(x) - f(y) > \varepsilon$ for all $y \in U_x$, with the result that there exists a nontrivial nonnegative continuous function $\chi(x)$ with a support belonging to U_x and such that $f(x) \leqslant (\mathrm{cl}\, f)(x) - \varepsilon\chi(x) < (\mathrm{cl}\, f)(x)$ at points where $\chi(x) \neq 0$. The existence of such a function contradicts the definition (1.1.3.6).

Let us prove the validity of (1.1.3.7). Since $\mathrm{cl}\, f \geqslant f$, it is clear that the right-hand side of (1.1.3.7) is no smaller than the left-hand side. Suppose that $\theta = \sup (f + \varphi) < \sup (\mathrm{cl}\, f + \varphi) = \Xi = \theta + \sigma$, with $\sigma > 0$. According to the definition of an upper bound for an arbitrarily small positive ε there is always a point x_ε such that $(\mathrm{cl}\, f)(x_\varepsilon) + \varphi(x_\varepsilon) > \Xi - \varepsilon = \theta + \sigma - \varepsilon$. Since $\varphi(x)$ is semi-continuous below, the set $U^\varepsilon = \{x \colon \varphi(x) > \varphi(x_\varepsilon) - \varepsilon\}$ is open and contains the point x_ε, that is, U^ε constitutes an open neighborhood of point x_ε. In view of the remark made in the proof of the theorem, on set U^ε there exists a point y such that $(\mathrm{cl}\, f)(x_\varepsilon) - f(y) \leqslant \varepsilon$, whereby $f(y) - \varphi(y) \geqslant (\mathrm{cl}\, f)(x_\varepsilon) + \varphi(x_\varepsilon) - 2\varepsilon > \theta + \sigma - 3\varepsilon$.

Selecting now ε smaller than 3σ, we arrive at a contradiction with the definition $\theta = \sup_x (f + \varphi)$, which implies that $\theta = \Xi$. The proof of the theorem is complete.

Reasoning along the same lines, we can also prove that

$$\inf_x (f(x) + \varphi(x)) = \inf_x ((\mathrm{cl}_* \varphi)(x) + f(x)), \tag{1.1.3.8}$$

where f is a semi-continuous-above function, and cl_* stands for the lower closure

$$(\mathrm{cl}_* \varphi)(x) = \sup \{\psi(x) : \psi \leqslant \varphi, \quad \psi \in C(R^n)\}. \quad (1.1.3.9)$$

For every semi-continuous-above function ψ we have

$$(\mathrm{cl} \circ \mathrm{cl}_* \psi)(x) \leqslant \psi(x),$$

while for every semi-continuous-below function η we have $\mathrm{cl}_* \circ \mathrm{cl}\, \eta \geqslant \eta$. Whence $\sup\limits_x (\eta' + \mathrm{cl}_* \circ \mathrm{cl}\, \eta'') \leqslant \sup\limits_x (\eta' + \mathrm{cl}\, \eta'')$ and $\inf\limits_x (\mathrm{cl} \circ \mathrm{cl}_* \psi' + \psi'') \geqslant \inf\limits_x (\mathrm{cl}_* \psi' + \psi'')$. Combining this with (1.1.3.7) and (1.1.3.8), we get

$$\sup_x (\eta' + \eta'') = \sup_x (\eta' + \mathrm{cl}_* \circ \mathrm{cl}\, \eta'') = \sup_x (\eta' + \mathrm{cl}\, \eta''),$$
$$(1.1.3.10)$$
$$\inf_x (\psi' + \psi'') = \inf_x (\mathrm{cl} \circ \mathrm{cl}_* \psi' + \psi'') = \inf_x (\mathrm{cl}_* \psi' + \psi'').$$

For the domain of functions f and φ we can take any normal topological space, since such a space allows for the existence of nontrivial continuous real functions whose support lies within an open neighborhood of any given point. This is the only topological property of the domain of functions employed in the proof of (1.1.3.7)-(1.1.3.10).

In what follows we describe the procedure of constructing the maximal σ-additive idempotent continuation of a measure from algebra Σ to the σ-algebra generated by Σ and prove the following

Theorem 1.1.3.2 *Let $f(x)$: $R^n \to R$ be a real function bounded above, Σ an algebra generating the σ-algebra $\mathscr{B}(R^n)$, and $m(B) = \sup\limits_{x \in B} f(x)$, $B \in \Sigma$. Then there exists a maximal σ-additive idempotent continuation m^* of measure m from Σ to the σ-algebra $\mathscr{B}(R^n)$, with $m^*(\beta) = \sup\limits_{x \in B} (\mathrm{cl}\, f)(x)$, $\beta \in \mathscr{B}(R^n)$.*

Theorems 1.1.3.1 and 1.1.3.2 show that if f is a bounded real function and φ is a bounded semi-continuous-below function, then the upper bound $\sup\limits_x (f + \varphi)$ can be interpreted as the idempotent Lebesgue integral

$$I(\varphi) = \int_{R^n}^{\oplus} \varphi(x) \odot m(dx), \quad (1.1.3.11)$$

where m is any σ-additive idempotent continuation of measure $m(B) = \sup\limits_{x \in B} f(x)$, $B \in \Sigma$, with the measure given on any algebra Σ generating the σ-algebra $\mathscr{B}(R^n)$. Here the right-hand side of (1.1.3.11) does not depend on the choice of continuation.

Thus, the property of semi-continuity (below) of an integrable function compensates for the absence of continuity of an idempotent measure on empty sets and the related existence of distinct continuations of the measure. In this sense the semi-continuity of functions is dual to the continuity of measures on empty sets.

Let us now answer a question that emerges when one considers the integral $I(\varphi)$, namely, for what functions φ is the integral $I(\varphi)$ independent of the way one continues the measure from algebra Σ when the topology of the domain of φ is not known (and hence the continuity of φ is not defined)? Below we show that such functions must possess the following property:

$$\{x\colon \varphi(x) > c\} \in G, \tag{1.1.3.12}$$

where G is the class of sets generated by countable unions of sets belonging to Σ. By analogy with the definition of functions semi-continuous below, we will call such functions semi-measurable below.

We will now consider several examples showing that the choice of measure and algebra Σ essentially determines the complexity of the numerical estimate of the upper bound.

If f and φ are two real continuous-below bounded-above functions on R^n, then $I = \sup\limits_x (f + \varphi)$ can be represented in the form of two distinct idempotent integrals:

$$I = \int^{\oplus} f(x) \odot m_\varphi(dx) = \int^{\oplus} \varphi(x) \odot m_f(dx),$$

where $m_\varphi(B) = \sup\limits_B \varphi(x)$, and $m_f(B) = \sup\limits_B f(x)$. This implies the possibility of various algorithms for estimating $\sup\limits_B (f + \varphi)$. In one case to estimate the integral one must calculate $\sup\limits_{Q_i^\varepsilon} (x)$ on the sets $Q_i^\varepsilon = \{x\colon f(x) > d_i^\varepsilon\}$, where $\{d_i^\varepsilon\}$ is an ε-mesh on the set of values of map f, and then estimate sup by sums $\oplus$: $I \simeq \bigoplus\limits_i (d_i^\varepsilon \odot m_\varphi(Q_i^\varepsilon))$. In the other $I \simeq \bigoplus\limits_i (g_i^\varepsilon \odot m_f(D_i^\varepsilon))$, where $D_i^\varepsilon = \{x\colon \varphi(x) > g_i^\varepsilon\}$, and $\{g_i^\varepsilon\}$ is an ε-mesh on the set of values of map φ.

The difference between these two methods of estimation proves to be important when one of the functions, say φ, is simple (see below) and assumes a finite number of values; the use of such a function as the integrand makes the second estimate exact for all ε sufficiently small and, hence, simplifies estimation of the integral.

It is important to note at this stage that computers are unable in principle to carry out an exact calculation of $\sup \varphi(x)$ on an arbitrary set provided by the σ-algebra $\mathfrak{A}$. Even if φ is monotonic, $\sup \varphi(x)$

can be calculated only on segments with rational end points, that is, on sets provided by a subalgebra of Σ_R. Theorem 1.1.3.2 shows that the estimates of sup $(f + \varphi)$ obtained numerically by the Lebesgue method will converge to the value of sup $(f + \varphi)$ if $\varphi(x)$ is semi-continuous (semi-measurable) below.

The problem concerning the lack of dependence of the integral on the method of continuation of measure from algebra Σ_R to σ-algebra $\mathfrak{A}$ becomes especially important when the variable of integration assumes values in an infinitely dimensional space and the values of the measure can be given constructively only on Σ. Let us show that different algorithms of the estimation of the integral correspond to different variants of the algebra that generates the σ-algebra $\mathfrak{A}$.

Suppose that $\Omega = \{x_\tau\colon [0, t] \to R,\ x_0 = 0,\ \int_0^t |\dot{x}_\tau|^2 d\tau < \infty\}$, and the algebra $\Sigma = \Sigma_C$ consists of so-called cylinder sets

$$S = S(\xi_\tau^1, \ldots, \xi_\tau^k | B_k) = \{x_\tau\colon (\langle \dot{x}_\tau, \xi_\tau^1\rangle, \ldots, \langle \dot{x}_\tau, \xi_\tau^k\rangle) \in B_k\}, \tag{1.1.3.13}$$

where $B_k \subset \mathscr{B}(R^k)$, $x_\tau \in \Omega$, $\xi_\tau^n \in C[0, t]$, and $\langle \dot{x}_\tau, \xi_\tau\rangle = \int_0^t \dot{x}_\tau \dot{\xi}_\tau\, d\tau$.

Suppose also that $m(S) = -\sup_{x_\tau \in S}\left\{-\int_0^t |\dot{x}_\tau|\, d\tau\right\}$ is the measure on Σ_C. Here is a method for estimating its values. Space Ω can be represented in the form of the direct sum of the finite-dimensional space $\mathscr{L}_k = \mathscr{L}(\xi_\tau^1, \ldots, \xi_\tau^k) = \left\{x_\tau^\lambda\colon \dot{x}_\tau = \sum_{j=1}^k \lambda_j \xi_\tau^j,\ x_0 = 0\right\}$ and its orthogonal complement $\mathscr{L}_k^\perp$. Clearly, if $x_\tau \in S_k$ and $y_\tau \in \mathscr{L}_k$, then $x_\tau + x_\tau \in S$ for every $y \in \mathscr{L}_k^\perp$. On the other hand,

$$\int_0^t |\dot{x}_\tau|^2 d\tau \leqslant \int_0^t |\dot{y}_\tau + \dot{x}_\tau|^2 d\tau = \int_0^t (|\dot{x}_\tau|^2 + |\dot{y}_\tau|^2)\, d\tau.$$

Whence, the upper bound on S coincides with the upper bound on a subset of the finite-dimensional space $\mathscr{L}_k$:

$$m(S) = \inf_{\substack{x_i(\lambda) \in S \\ \lambda \in R^k}} \int_0^t |\dot{x}_\tau(\lambda)|^2 d\tau. \tag{1.1.3.14}$$

Calculation of the lower bound of (1.1.3.14) on subsets of Σ_C constitutes a quadratic-programming problem. Thus, the estimate

of measure $m(S)$, $S \in \Sigma_C$, can be obtained numerically. In addition to the algebra Σ_C of cylinder sets, let us introduce an algebra Σ_M generated by sets of the type

$$S = S(t_1, \ldots, t_k | B_k) = \{x_\tau : (x_{t_1}, \ldots, x_{t_k}) \in B_k\},$$
$$0 < t_1 < \ldots < t_k < t.$$

Determining the measure $m(S)$ on sets of this type can be reduced to a simpler quadratic-programming problem, whose solution does not require scalar products of the type $\langle \xi_\tau^n, \xi_\tau^j \rangle = \int_0^t \xi_\tau^n \xi_\tau^j \, d\tau$:

$$m(S) = \inf_{(x_1, \ldots, x_k) \in B_k} \left\{ \inf_{\substack{x_\tau : x_{t_j} = x_j \\ x_0 = 0}} \int_0^t |\dot{x}_\tau|^2 \, d\tau \right\}$$
$$= \inf_{(x_1, \ldots, x_k) \in B_k} \sum_{j=0}^{k-1} |x_{j+1} - x_j|^2/(t_{j+1} - t_j), \quad x_0 = 0. \tag{1.1.3.15}$$

The right-hand side of (1.1.3.15) is the lower bound of a quadratic function given in the form of an analytical expression. Convergence of the estimate algorithms for idempotent integrals by virtue of measure values on the algebras Σ_C and Σ_M is a corollary of analogs of the Lebesgue theorem on the passage to the limit under the integral sign.

To describe the weak convergence of operators, we topologize the set of semi-continuous-below functions. Let A be a partially ordered metric semi-ring. To ensure the continuity of the A-valued scalar product $(f, \varphi) = \sup_{x \in X} (f(x) \dotplus \varphi(x))$, we match the topology on the set of semi-continuous-below functions with the topology on semi-ring A. For the topologies of this type we take the topologies generated by various metrics ρ on A

$$r(\varphi_1, \varphi_2) = \sup_x \rho(\varphi_1(x), \varphi_2(x)), \tag{1.1.3.16}$$

say

$$r_{\tan^{-1}}(\varphi_1, \varphi_2) = \sup_x [\tan^{-1}(\max\{\varphi_1(x), \varphi_2(x)\})$$
$$- \tan^{-1}(\min\{\varphi_1(x), \varphi_2(x)\})].$$

The set of functions semi-continuous-below (above) is complete in the nonuniform-convergence topology, and the continuous map of a function semi-continuous-below (above) is also semi-continuous-below (above). These facts can be employed to prove that the space of semi-continuous-below functions in metrics $r_{\exp}$ and $r_{\tan^{-1}}$ is

dense. Suppose that $\{\varphi_n^1\}_1^\infty$ and $\{\varphi_n^2\}_2^\infty$ are sequencies of semi-continuous-below functions; these sequences are assumed to be Cauchy sequences in the metrics $r_{\exp}$ and $r_{\tan^{-1}}$, respectively. Then the sequences $\{\exp\{\varphi_n^1(x)\}\}$ and $\{\tan^{-1}\{\varphi_n^2(x)\}\}$ also consist of functions that are semi-continuous-below, with

$$|\exp\{\varphi_n^1(x)\} - \exp\{\varphi_m^1(x)\}| \leqslant r_{\exp}(\varphi_n^1, \varphi_m^1),$$
$$|\tan^{-1}\{\varphi_n^2(x)\} - \tan^{-1}\{\varphi_m^2(x)\}| \leqslant r_{\tan^{-1}}(\varphi_n^2, \varphi_m^2).$$

Thus, the sequences $\{\exp\{\varphi_n^1(x)\}\}$ and $\tan^{-1}\{\varphi_n^2(x)\}$ are also Cauchy sequences in the uniform-convergence metric and their limits,

$$\varphi_{\exp}(x) = \lim_{n\to\infty}\{\exp\{\varphi_n^1(x)\}\} \quad \text{and} \quad \varphi_{\tan^{-1}}(x) = \lim_{n\to\infty}\{\tan^{-1}\{\varphi_n^2(x)\}\},$$

are semi-continuous below. The inverse functions $\ln(\varphi_{\exp}(x))$ and $\tan(\varphi_{\tan^{-1}}(x))$ are semi-continuous below, too. This implies that the considered metric spaces of functions continuous below are complete.

The metrics $r_{\exp}$ and $r_{\tan^{-1}}$ examined here possess three important properties:

(a) Uniformity with respect to semi-group operations on a semi-ring: if $R(a_n, b_n) \to 0$ as $n \to \infty$, then $R(a_n \diamondsuit d, b_n \diamondsuit d)$ tends to 0 as $n \to \infty$ uniformly with respect to $d \in \mathscr{D}$ for every bounded set $\mathscr{D}$, where R is one of the metrics, and $\diamondsuit$ is one of the semi-group operations, that is, $\oplus$ or $\odot$ or max or min.

(b) The minimax property:

$$R((a \diamondsuit b), (c \diamondsuit d)) \leqslant \max\{R(a, c), R(b, d)\},$$

where R is one of the metrics, and $\diamondsuit$ is one of the three semi-group operations $\odot$ or max or min. This property implies the minimax inequality

$$R(\sup_{i\in I}\{a_i\}, \sup_{j\in J}\{b_j\}) \leqslant \inf_{i\in I} \sup_{j\in J} R(a_i, b_i)$$

for all sets I and J.

(c) Monotonicity: if $a \leqslant b \leqslant c$, then

$$\max\{\rho(a, b), \rho(b, c)\} \leqslant \rho(a, c).$$

The conditions sufficient for the continuity of A-valued functional are formulated in the following

Theorem 1.1.3.3 *If a metric r on the set of A-valued functions $\varphi: X \to A$ is related to a metric of the semi-ring A through* (1.1.3.16) *and possesses the property of uniformity and the minimax property, then the A-valued linear functional on a metric space of A-valued func-*

tions that acts according to the rule $(f, \varphi) = \sup\limits_X (f \diamondsuit \varphi(x))$ *is continuous in the metric of A for every bounded function f.*

Proof. Suppose we have defined the limit of $r(f, \varphi_n)$ as $n \to \infty$ and suppose that f is a bounded function: $r(f, \mathbb{0}) < \infty$. Then, by virtue of the properties mentioned in the hypothesis, we get

$$\begin{aligned} \rho_n &= \rho((f, \varphi_n), (f, \varphi)) = \rho(\sup_X (\varphi_n \diamondsuit f)(x), \sup_X (\varphi \diamondsuit f)(x)) \\ &\leqslant \sup_X (\rho(\varphi_n \diamondsuit f)(x), (\varphi \diamondsuit f)(x)) \\ &= r((\varphi_n \diamondsuit f), (\varphi \diamondsuit f)) = r_n, \end{aligned}$$

which implies that ρ_n tends to zero (together with r_n) as $n \to \infty$. The proof is complete.

Now let $\mathcal{L}(h)$, $h \in [0, 1]$, be a family of continuous operators acting in a metric space of A-valued functions, while $\mathcal{L}^*(h)$ is a family of conjugate operators defined via the A-valued scalar product

$$\sup_{x \in X} (\mathcal{L}(h) f(x) \diamondsuit \varphi(x)) = \sup_{x \in X} (f(x) \diamondsuit \mathcal{L}^*(h) \varphi(x)).$$

If $\mathcal{L}^*(h)$ is continuous in h in metric r, there exists a limit $\lim\limits_{h \to 0} \mathcal{L}^*(h) \times \varphi(x) = (\mathcal{L}^* \varphi)(x)$ that defines an A-valued continuous linear functional f^*

$$(f^*, \varphi) = \sup_X (f \diamondsuit \mathcal{L}^* \varphi).$$

In particular, if $\diamondsuit = \oplus$, then, in accordance with Theorem 1.1.3.3, the weak limit f^* coincides, to within an isomorphism, with the equivalence class of functions with the same closure and can be determined uniquely by its values on the set of semi-continuous-below functions that assume the values $-\infty = \mathbb{0}$ and $0 = \mathbb{1}$. If $X = R^n$, then for the subset of the main space that weakly separates the set of continuous A-valued functionals we can use the nonnegative functions from $C^\infty(R^n)$.

We now wish to prove the theorem on weak separability with respect to the scalar product $(f, \varphi) = \sup\limits_X (f(x) + \varphi(x))$. Suppose that B is an arbitrary set of functions $\varphi: X \to R$. We call $P_B f(x) = \inf\limits_{\varphi \in B} ((f, \varphi^*) + \varphi(x))$, with $\varphi^*(x) = -\varphi(x)$, the projection of function $f: X \to R$ with respect to B. We then have

Theorem 1.1.3.4 $(f_1, \varphi) = (f_2, \varphi)\ \forall \varphi \in B$ *if and only if* $P_B f_1 = P_B f_2$.

Proof. Let us see whether $(P_B f, \varphi) = (f, \varphi)$ $\forall \varphi \in B$. Clearly, $P_B f \geqslant f$ and $(P_B f, \varphi) \geqslant (f, \varphi)$. On the other hand, $(P_B f, \varphi) \leqslant \sup\limits_X ((f, \varphi) - \varphi(x) + \varphi(x)) = (f, \varphi)$. Whence, $(P_B f, \varphi) = (f, \varphi)$. Sufficiency has been proved.

Necessity will be proved by contradiction. Suppose that $P_B f_1 \not\equiv P_B f_2$. Then there exists a point x_c, $c > 0$, such that $P_B f_1(x_c) \geqslant P_B f_2(x_c) + c$. By the definition of P_B, for every positive ε there exists a function $\psi = \psi_{\varepsilon, c} \in B$ such that $P_B f_2(x_c) + \varepsilon \geqslant (f_2, \psi^*) + \psi(x_c)$. Whence, $(f, \psi^*) + \psi(x_c) \geqslant P_B f_1(x_c) \geqslant P_B f_2(x_c) + c \geqslant (f_2, \psi^*) + \psi(x_c) + c - \varepsilon$.
This implies that $(f_1, \psi^*) \geqslant (f_2, \psi^*) + c - \varepsilon$, which contradicts the hypothesis for $\varepsilon < c$. From this we conclude that $P_B f_1 \equiv P_B f_2$. The proof of the theorem is complete.

Now we wish to formulate a theorem on the weak upper bound on a sequence of linear idempotent functionals. Let X be a locally compact topological space. $C_0(X)$ a set of continuous real functions on X with a compact support, and $\{f_n\}_1^\infty$ a sequence of functions on X with values on the extended real axis $R \cup \{\pm\infty\}$. By $C_B(X)$ we denote the set of all semi-continuous-below real functions on X with values in $R \cup \{\pm\infty\}$.

For the upper envelope $\Phi_B(x)$ of the sequence $\{f_n\}_1^\infty$ with respect to set B we take the lower bound of the set of all the semi-continuous-above functions that are no smaller than all the functions of the sequence $\{P_B f_n(x)\}_1^\infty$ starting from a certain function:

$$\Phi_{B,n}(x) = \inf\{\Phi(x)\colon\ \Phi(\xi) \geqslant P_B f_k(\xi) \quad \forall \xi \in X, \quad \forall k \geqslant n, \ \Phi \in C_B\},$$
$$\Phi_B(x) = \inf_n \Phi_{B,n}(x) = \lim_{n\to\infty} \Phi_n(x). \tag{1.1.3.17}$$

Remark 1.1.3.1 Since $\Phi_{B,n} \downarrow \Phi_B$ as $n \to \infty$ and since the $\Phi_{B,n}$ are semi-continuous above, the function Φ_B is semi-continuous above, too.

Before we go over to the main theorem on weak convergence, let us consider an example that illustrates the essence of this theorem.

Let $X = [0, 1]$ and $f_n(x) = a(x) \cos nx$, where $a(x)$ is a continuous bound function. Then $\Phi_B(x) = |a(x)|$ and there exists a weak limit $\lim_{n\to\infty} (f_n, \varphi) = (\Phi_B, \varphi)$, that is, $|a(x)|$ is the weak limit of the sequence of functions $f_n(x) = a(x) \cos nx$ with respect to the sup-sum scalar product.

Indeed, clearly $\Phi_B(x) \geqslant |f_n(x)| \geqslant f_n(x)$, whereby

$$\overline{\lim_{n\to\infty}} (f_n, \varphi) \leqslant (\Phi_B, \varphi). \tag{1.1.3.18}$$

Suppose now that $\varepsilon > 0$ and that $x_\varepsilon \in [0, 1]$ is a point at which $(\Phi_B, \varphi) \leqslant \Phi_B(x_\varepsilon) + \varphi(x_\varepsilon) + \varepsilon$. For a function $\varphi(x)$ that is semi-continuous below, the set $M_\varepsilon = \{x\colon \varphi(x) > \varphi(x_\varepsilon) - \varepsilon\}$ is open

and contains point x_ε. Where there are sufficiently large values of n, the set M_ε also contains the point y_ε that is no farther from point x_ε than by $2\pi/n$ and at which $\cos(ny_\varepsilon) = \operatorname{sgn} a(y_\varepsilon)$ and $f_n(y_\varepsilon) = |a(y_\varepsilon)| = \Phi_B(y_\varepsilon)$. Since $a(x)$ is a continuous function, the set

$$N_\varepsilon = \{x\colon\ \Phi_B(x) < \Phi_B(y_\varepsilon) + \varepsilon\}$$

is open and contains point y_ε as well as point x_ε for all sufficiently large values of n.

Thus, x_ε, $y_\varepsilon \in M_\varepsilon \cap N_\varepsilon$ and

$$\begin{aligned}(f_n, \varphi) &\geqslant f_n(y_\varepsilon) + \varphi_n(y_\varepsilon)\\ &\geqslant \Phi_B(x_\varepsilon) + \varphi(x_\varepsilon) - 2\varepsilon \geqslant (\Phi_B, \varphi) - 3\varepsilon.\end{aligned} \tag{1.1.3.19}$$

Since ε can be as small as desired, (1.1.3.19) yields the lower bound

$$\varliminf_{n\to\infty} (f_n, \varphi) \geqslant (\Phi_B, \varphi). \tag{1.1.3.20}$$

Now (1.1.3.18)-(1.1.3.20) imply $\lim\limits_{n\to\infty} (f_n, \varphi) = (\Phi_B, \varphi)$. In general we have the following

Theorem 1.1.3.5 *For every continuous real function $\varphi(x)$ with a compact support and every sequence of real functions, $\{f_n(x)\}_1^\infty$, there exists an upper bound*

$$\begin{aligned}\varlimsup_{n\to\infty} \sup_X (f_n(x) + \varphi(x)) &= \sup_X (\Phi_B(x) + \varphi(x))\\ &= (\Phi_B, \varphi),\end{aligned} \tag{1.1.3.21}$$

where Φ_B is the upper envelope (1.1.3.17) *of the sequence $\{f_n\}$.*

Proof. In the first place, using the fact that the support of φ is compact and that φ and $\Phi_{B,n}$ are semi-continuous above, we arrive at the following upper bound: $\varlimsup\limits_{n\to\infty} (f_n, \varphi) \leqslant (\Phi_B, \varphi)$. Then, employing the fact that Φ_B is semi-continuous above and φ is semi-continuous below, we arrive at the lower bound: $\varliminf\limits_{n\to\infty} (f_n, \varphi) \geqslant (\Phi_B, \varphi)$ separately for $(\Phi_B, \varphi) < +\infty$ and for $(\Phi_B, \varphi) = +\infty$. These bounds comprise the assertion of the theorem, while the requirements imposed on φ imply, in toto, the continuity of φ.

The inequalities $f_n(x) \leqslant \Phi_{B,n}(x)$ and $(f_n, \varphi) \leqslant (\Phi_{B,n}, \varphi)$, which follow from the definition of an upper envelope, show that the derivation of the upper bound is reduced to checking the inequality $\lim\limits_{n\to\infty} (\Phi_{B,n}, \varphi) \leqslant (\Phi_B, \varphi)$. Let us verify the assumption that $(\Phi_{B,n}, \varphi) > (\Phi_B, \varphi) + \delta$, $\delta > 0$, contradicts the property of semi-continuity above.

Indeed, in this case on the compact set $K = \operatorname{supp} \varphi$ there exists a sequence of points $\{x_{n,\delta}\}_1^\infty$ such that

$$\Phi_{B,n}(x_{n,\delta}) + \varphi(x_{n,\delta}) > (\Phi_B, \varphi) + \delta/2, \quad x_{n,\delta} \in K.$$

By virtue of the compactness of set K, from the sequence $\{x_{n,\delta}\}_1^\infty$ we can select a converging subsequence $\{x_{n_k,\delta}\} \to y^* \in K$ as $k \to \infty$. We denote $x_{n_k,\delta}$ by y_k. Thus, every open neighborhood U of point y^* contains the points y_k, $k \geqslant k_0(U, \delta)$, at which $\Phi_{B,n_k}(y_k) + \varphi(y_k) > (\Phi_B, \varphi) + \delta/2$. The subsequence Φ_{B,n_k} is not monotone increasing, so that $\Phi_{B,n}(y_k) + \varphi(y_k) \geqslant (\Phi_B, \varphi) + \delta/2$ for all $n \leqslant n_k$. By virtue of the fact that $\varphi(y)$ is semi-continuous above, we can always select a $k = k_0(\delta)$ so large that $\varphi(y_k) \leqslant \varphi(y^*) + \delta/4$, with

$$\Phi_{B,n}(y_k) + \varphi(y^*) \geqslant (\Phi_B, \varphi) + \delta/4, \quad k \geqslant k_0(\delta),$$
$$n \leqslant n_k.$$

Since $\Phi_{B,n}$ is semi-continuous above, we obtain $\Phi_{B,n}(y^*) \geqslant \lim\limits_{k\to\infty} \Phi_{B,n}(y_k)$. This implies that $\Phi_{B,n}(y^*) + \varphi(y^*) = \lim\limits_{n\to\infty} \Phi_{B,n}(y^*) + \varphi(y^*) \geqslant (\Phi_B, \varphi) + \delta/4$, which contradicts the definition of a sup-sum scalar product. Thus, $\lim\limits_{n\to\infty} (\Phi_{B,n}, \varphi) \leqslant (\Phi_B, \varphi)$ and

$$\overline{\lim_{n\to\infty}} (f_n, \varphi) \leqslant (\Phi_B, \varphi). \tag{1.1.3.22}$$

Suppose that $(\Phi_B, \varphi) < \infty$. Then for every positive ε there is a point $x_\varepsilon \in X$ such that

$$(\Phi_B, \varphi) \leqslant \Phi_B(x_\varepsilon) + \varphi(x_\varepsilon) + \varepsilon. \tag{1.1.3.23}$$

The set $M_\varepsilon = \{x\colon \varphi(x) > \varphi(x) - \varepsilon\}$ is open and contains point x_ε. Let us assume that, for a fixed positive δ,

$$P_B f_n(y) \leqslant \Phi_B(x_\varepsilon) - \delta \quad \forall y \in M_\varepsilon \tag{1.1.3.24}$$

for all n's starting from a definite one. Then the function

$$\Phi(x) = \{\infty,\ x \in X \setminus M_\varepsilon,\ \Phi_B(x_\varepsilon) - \delta,\ x \in M_\varepsilon\} \in C_B(X)$$

is semi-continuous above and $\Phi(x) \geqslant P_B f_n(x) \quad \forall x \in X$. The definition (1.1.3.17) on an upper envelope and the assumption (1.1.3.24) lead to a contradiction:

$$\Phi_B(x_\varepsilon) \leqslant \Phi(x_\varepsilon) = \Phi_B(x_\varepsilon) - \delta.$$

Thus, (1.1.3.24) cannot be valid and for every positive δ and an infinitude of values of n there are points $y_{\delta,n} \in M_\varepsilon$ such that

$$P_B f_n(y_{\delta,n}) \geqslant \Phi_B(x_\varepsilon) - \delta, \quad y_{\delta,n} \in M_\varepsilon. \tag{1.1.3.25}$$

Combining Theorem 1.1.3.4, inequalities (1.1.3.23) and (1.1.3.25), and the definition of M_ε, we find that there is an infinite number of values of n for which

$$\begin{aligned}(f_n, \varphi) = (P_B f_n, \varphi) &\geqslant P_B f_n (y_{\delta,n}) + \varphi (y_{\delta, n}) \\ &\geqslant \Phi_B (x_\varepsilon) + \varphi (x_\varepsilon) - \varepsilon + \delta \\ &\geqslant (\Phi_B, \varphi) - 2\varepsilon - \delta.\end{aligned}$$

Since both ε and δ can be as small as desired, we have

$$\underline{\lim}_{n\to\infty} (f_n, \varphi) \geqslant (\Phi_B, \varphi), \tag{1.1.3.26}$$

and this combined with (1.1.3.22) yields the validity of the theorem.

The proof for $(\Phi_B, \varphi) = \infty$ is similar. For every positive N there is a point x_N such that $\Phi_B (x_N) + \varphi (x_N) \geqslant N$. Just as in the previous case, there exists a set $M_{N,\varepsilon} = \{x\colon \varphi (x) > \varphi (x_N) - \varepsilon\}$. This set is open, contains point x_N, and for an infinitude of values of n and any positive δ there are points $y_{n,\delta}$ such that $P_B f_n (y_{n,\delta}) \geqslant \Phi (x_N) - \delta$. Whence,

$$\begin{aligned}(f_n, \varphi) = (P_B f_n, \varphi) &\geqslant f_n (y_{n,\delta}) + \varphi (y_{n,\delta}) \\ &\geqslant \Phi (x_N) + \varphi (x_N) - \varepsilon - \delta \\ &\geqslant N - \varepsilon - \delta.\end{aligned}$$

Since ε and δ can be as small as desired and N can be as large, we have $\overline{\lim}_{n\to\infty} (f_n, \varphi) = \infty = (\Phi_B, \varphi)$. The proof of the theorem is complete.

Corollary *If the sequence $\{f_n\}_1^\infty$ is weakly convergent with respect to the sup-sum scalar product on the set $C_0(X)$ of continuous functions with a compact support, its weak limit is the upper envelope* (1.1.3.17). *For stationary sequences this result coincides with the assertion of Theorem* 1.3.4.

1.1.4 Maximal Continuation of an Idempotent Measure

In this section we will prove the constructive theorem on the maximal continuation of an idempotent measure. We will show that the difference between the minimal continuation and the maximal emerges when continuation on closed sets is considered.

Let Ω be a set, Σ the algebra of the subsets of this set, $\mathfrak{A}$ the smallest σ-algebra generated by Σ, and $\mu\colon \mathfrak{A} \to A$ an A-valued function of the set. We will assume that A is the partially ordered set that is an Abelian semi-group with respect to the associative operations $\oplus$ and $\odot$ with neutral elements $\mathbb{O}$ and $\mathbb{1}$ respectively. We assume, in addition, that A is a complete structure, with the operations sup and inf commutative with the semi-group operations

Let us consider an A-valued function of the set, μ: $\{\Sigma \setminus \{\varnothing\}\} \to A$, with the following properties:

(a) nonnegativity and boundedness, or

$$\mathbb{O} \leqslant \mu(S) \leqslant a, \quad \mu(\Omega) = a \in A, \quad S \in \Sigma;$$

(b) additivity, or

$$\mu(S \cup S') = \mu(S) \oplus \mu(S'), \quad S \cap S' = \varnothing, S, \ S' \in \Sigma;$$

(c) idempotency, or

$$\mu(S) \oplus \mu(S) = \mu(S).$$

By G we denote the class of subsets of Ω that are countable unions of sets belonging to Σ. We define the set function m: $G \to A$ thus:

$$m(g) = \sup_{\{S_n\}} \sup_n m\left(\bigcup_1^n S_i\right), \quad g \in G, \quad g = \bigcup_1^\infty S_i, \quad S_i \in \Sigma. \qquad (1.1.4.1)$$

We also wish to define the sets $m(a)$ belonging to $\mathfrak{A}$ that are nonempty countable intersections of sets belonging to G thus:

$$m(a) = \inf_{\{g_n\}} \inf_n m\left(\bigcap_1^n g_i\right), \quad g_i \in G, \quad a = \bigcap_1^\infty g_i \in \mathfrak{A}. \qquad (1.1.4.2)$$

Theorem 1.1.4.1 *The set function m is bounded, nonnegative, idempotent, and σ-additive on* $\mathfrak{A}$.

The idempotent σ-additive continuation of m is unique on G and maximal on $\mathfrak{A}$.

Proof. The boundedness and nonnegativity of m on G and $\mathfrak{A}$ are obvious. Let us prove the monotonicity of m on G. Suppose that $g' \supset g$. We take two sequences, $\{S'_n\}$ and $\{S_n\}$, such that $g' = \bigcup_1^\infty S'_n$ and $g = \bigcup_1^\infty S_n$. Then $\bigcup_1^n S'_j \supset \left(\bigcup_1^n S'_j\right) \cap \left(\bigcup_1^k S_i\right)$, and in view of the monotonicity of m on Σ for every pair (n, k) we have

$$m\left(\bigcup_1^n S'_j\right) \geqslant m\left(\left(\bigcup_1^n S'_j\right) \cap \left(\bigcup_1^k S_i\right)\right).$$

A similar inequality holds true also for the upper bounds in n:

$$m(g') \geqslant m\left(g' \cap \left(\bigcup_1^k S_i\right)\right) = m\left(\bigcup_1^k S_i\right).$$

Going over to upper bounds in k, we get $m(g') \geqslant m(g)$. The monotonicity of m on G has been proved.

We will now prove, in passing, that the values of $m(g)$ are independent of the choice of the sequence $\{S_i\}$. True, since if $g = \bigcup S_i \supseteq$

$\bigcup S'_i = g'$, we have, on the one hand, $m(g) \geqslant m(g')$ and, on the other, $m(g) \leqslant m(g')$. Whence, in definition (1.1.4.1), the upper bound in $\{S_n\}$ will not alter the result.

The monotonicity of m on $\mathfrak{A}$ can be proved by reasoning along similar lines. However, now in definition (1.1.4.2) we cannot discard the lower bound in $\{g_n\}$. Empty sets, which potentially can emerge in the passage to the limit of intersections, may transform $m(a)$, $a \in \mathfrak{A}$, into a constant and thus violate the properties of monotonicity, additivity, and idempotency.

For example, suppose that $\Sigma = \Sigma_R$ (see Section 1.1.3), $f(x) = \{x,\ x \in [0, 1];\ 0,\ x \notin [0, 1]\}$, and $m(B) = \sup\limits_{x \in B} f(x)$, $B \in \Sigma_R$. If $a = \bigcap\limits_1^\infty g_n$, $g_n \in G$, then $a = \bigcap\limits_1^\infty g'_n$, where $g'_n = g_n \cup (1 - 1/n,\ 1) \in G$. Without going over to the lower bound in $\{g_n\}$ we could have tried to determine the value of $m(a)$ for $a \in \mathfrak{A}$, $a \notin G$, in the following manner: $m(a) = \inf\limits_n m\left(\bigcap\limits_1^n g'_i\right)$. However,

$$m\left(\bigcap_1^n g'_i\right) = m(g'_n) = \sup\{x\colon\ x \in g_n \cup (1 - 1/n,\ 1\} \equiv 1 = m(R).$$

This example shows the important role played by the lower bound $\inf\limits_{\{g_n\}}$ in the definition (1.1.4.2).

Suppose that $a' \supset a$, a, $a' \in \mathfrak{A}$, and $a = \bigcap\limits_1^\infty g_i$ and $a' = \bigcap\limits_1^\infty g'_j$, g_i, $g'_j \in G$. Then $\bigcap\limits_1^n g'_i \supset \left(\bigcap\limits_1^n g'_i\right) \cap \left(\bigcap\limits_1^k g_j\right)$, and in view of the already proven monotonicity of m on G for all finite values of n and k we have

$$m\left(\bigcap_1^n g'_n\right) \geqslant m\left(\left(\bigcap_1^n g'_i\right) \cap \left(\bigcap_1^k g_j\right)\right) \geqslant m(a).$$

Going over to the lower bound in n and k, we find that $m(a') \geqslant m(a)$.

From the idempotency of m on Σ and the additivity on nonintersecting sets we conclude that m is additive on all sets belonging to Σ. Combining this result with the fact that $m(g \cup g')$ does not depend on the selection of the sequence from Σ that converges to $g \cup g'$, we conclude that m is additive on G:

$$m(g \cup g') = \sup_n \left(m\left(\left(\bigcup_1^n S_i\right) \cup \left(\bigcup_1^n S'_j\right)\right)\right)$$

$$= \sup_n \left(m\left(\bigcup_1^n S_i\right) \oplus m\left(\bigcup_1^n S'_j\right)\right) = m(g) \oplus m(g').$$

From the monotonicity of m on $\mathfrak{A}$ and idempotency of addition $\oplus$ it follows that

$$m(a \cup a') = m(a \cup a') \oplus m(a \cup a') \geqslant m(a) + m(a'). \quad (1.1.4.3)$$

On the other hand, the definition of m on $\mathfrak{A}$ implies the opposite inequality

$$m(a \cup a') \leqslant \inf_{\{g_n\},\{g'_n\}} \inf_n m\left(\left(\bigcap_1^n g_i\right) \cup \left(\bigcap_1^n g'_j\right)\right)$$
$$= m(a) \oplus m(a'). \quad (1.1.4.4)$$

The additivity and idempotency of m on $\mathfrak{A}$ follows from (1.1.4.3) and (1.1.4.4).

We now note that, on the one hand, for every finite n we have

$$m\left(\bigcup_1^n g_i\right) = \bigoplus_1^n m(g_i) \leqslant \bigoplus_1^\infty m(g_i),$$

whereby $m\left(\bigcup_1^\infty g_i\right) \leqslant \bigoplus_1^\infty m(g_i)$, while on the other, by virtue of the monotonicity of m on $\mathfrak{A}$, we have

$$m\left(\bigcup_1^\infty g_i\right) \geqslant m\left(\bigcup_1^n g_i\right) \geqslant \bigoplus_1^n m(g_i),$$

whereby $m\left(\bigcup_1^\infty g_i\right) \geqslant \bigoplus_1^\infty m(g_i)$. This implies the σ-additivity of m on G. The σ-additivity of m on $\mathfrak{A}$ can be proved in a similar manner.

Any idempotent σ-additive continuation of μ on G coincides with m. Indeed, suppose that w is some other continuation. Then

$$w(g) = \bigoplus_1^\infty w(S_i) = \bigoplus_1^\infty m(S_i) = m(g), \quad S_i \in \Sigma.$$

Let us demonstrate that any idempotent σ-additive continuation of μ on $\mathfrak{A}$ does not exceed m. Suppose that v is such a continuation. By virtue of its idempotency and σ-additivity, v coincides with m on G. The monotonicity of v implies $v(a) \leqslant v(g) = m(g)$ for every $g \supset a$, $g \in G$. There is a similar inequality for inf in g:

$$v(a) \leqslant \inf_n m\left(\bigcap_1^n g_i\right) = m(a), \quad \bigcap_1^\infty g_i = a.$$

Thus, the continuation of μ on $\mathfrak{A}$ constructed here is maximal. The proof of the theorem is complete.

The examples considered above show that in general smaller continuations are possible.

The theorem that we now formulate shows that A-valued idempotent measures corresponding to functions semi-continuous above coincide with their maximal continuations.

Theorem 1.1.4.2 *Suppose that* Ω *is a topological space,* $\mathscr{B} = \mathscr{B}(\Omega)$ *is the* σ*-algebra of Borel subsets, and* Σ *is an algebra that gener-*

ates $\mathscr{B}$. If f is an A-valued function semi-continuous above, then $\sup_{\omega\in a} f(\omega) = m(a)$, $a \in \mathscr{B}$, *where m is the maximal continuation of the idempotent measure* $\mu(S) = \sup_{\omega\in S} f(\omega)$, $S \in \Sigma$.

Proof. Clearly, the restriction m to class G coincides with the continuation of μ on G, since such a continuation is unique (see Theorem 1.1.4.1). On the σ-algebra $\mathscr{B}$ the measure m is maximal, whereby

$$m(a) \geqslant \sup_a f(x), \quad a \in \mathscr{B}.$$

Let us assume that there exists a set $a \in \mathscr{B}$ such that $\Xi = m(a) > \sup_a f(x) = \theta$. By virtue of the semi-continuity of $f(x)$ above, $\sup_a f(x) = \sup_{\bar a} f(x)$, where $\bar a$ is the closure of set a in the topology Ω. Suppose, in addition, that

$$\Xi = m(a) \overset{\text{def}}{=} \inf_{\{g_i\}} \inf_i m(g_i) \leqslant \inf_i m(\bar g_i) = \inf_i \sup_{x\in \bar g_i} f(x),$$

where $a = \bigcap_1^\infty g_i$.

Let us consider the set $\bar g = \{x \colon f(x) \geqslant \Xi\}$, where $\Xi \geqslant \theta$. By virtue of the semi-continuity of f above, set $\bar g$ is closed, whereby the sets $G_i = \bar g_i \cap \bar g \subset \bar g_i$ and their intersection $G_0 = \bigcap_1^\infty G_i \subset \bar a$ are closed and nonempty. Hence,

$$\Xi \leqslant \sup_{x\in G_0} f(x) \leqslant \sup_{x \in \bar a} f(x) = \theta,$$

which implies that $\theta = \Xi$. The proof of the theorem is complete. The assertion of this theorem also constitutes the assertion of Theorem 1.1.3.2 as a particular case.

1.1.5 Idempotent Integration of Measurable and Semi-measurable Functions

We will prove the theorem that the integral of a function that is semi-measurable below is independent of the choice of the continuation of the idempotent measure. Thus, the fact that the function is semi-measurable below compensates for the absence of continuity of the measure at zero.

Let Σ be an algebra of the subsets of Ω that generates the σ-algebra $\mathfrak{A}$ and let G be a class closed with respect to a countable union of elements from Σ. Suppose that A is a partially ordered metric semiring whose metric and partial ordering obey the conditions discussed in Section 1.1.2.

A function $f\colon \Omega \to A$ is said to be measurable below if the sets $V(a) = \{\omega\colon f(\omega) \geqslant a\}$ and $\Omega(a) = \{\omega\colon f(\omega) > a\}$ belong to $\mathfrak{A}$

for every $a \in A$, and semi-measurable below if $\Omega(a) \subset G$. The ordering relation on the semi-ring A is partial, whereby the set $W(a) = \{\omega: f(\omega) \leqslant a\}$ is not a complement of $\Omega(a)$ and we must distinguish between (semi)measurability above and (semi)measurability below. Naturally, the concept of semi-measurability depends on the choice of the algebra Σ that defines class G. However, in what follows we speak only of a fixed algebra Σ, whereby it becomes unnecessary to specify with respect to which class semi-measurability is defined.

Let us give a more precise definition of an ε-mesh. A subset $\{d_i^\varepsilon\}_1^\infty$ of set $\mathscr{D} \subset A$ is said to be an ε-mesh if for each point $d \in \mathscr{D}$ there exists an element d_i^ε such that

$$\rho(d_i^\varepsilon, d) \leqslant \varepsilon. \tag{1.1.5.1}$$

If in addition to (1.1.5.1) we also require that $d_i^\varepsilon \geqslant d$, the ε-mesh is said to be upper, while for $d_i^\varepsilon \leqslant d$ the ε-mesh is said to be lower. Note that although an arbitrary ε-mesh may be neither an upper one nor a lower one, to each ε-mesh there corresponds an upper ε-mesh and a lower 2ε-mesh:

$$\mathscr{D}_i^{2\varepsilon} = \sup \{\mathscr{D}: \rho(\mathscr{D}, d_i^\varepsilon) \leqslant \varepsilon\}$$

and

$$g_i^{2\varepsilon} = \inf \{g: \rho(g, d_i^\varepsilon) \leqslant \varepsilon\}.$$

Indeed, in view of the triangle property,

$$\rho(x, \mathscr{D}_i^{2\varepsilon}) \leqslant \rho(x, d_i^\varepsilon) + \rho(d_i^\varepsilon, \mathscr{D}_i^{2\varepsilon}) \leqslant 2\varepsilon.$$

The function $f: \Omega \to A$ is said to be elementary if it assumes the value $f =$ const on a set $a \in \mathfrak{A}$ and $\mathbb{O}$ on its complement. A function f equal to no more than the countable sum of elementary functions is said to be simple: $f_s(\omega) = \bigoplus\limits_1^\infty f_i(\omega)$.

To define the concept of an A-valued idempotent integral, we will employ some properties of the metric and structure of the semi-ring A. These are formulated below in the form of conditions:

(1) The matching of metric and structure: if $\sup a_n \to a$ and $\inf a_n \to a$ as $n \to \infty$, then $\rho(a_n, a) \to 0$, and vice versa.

(2) The uniformity of metric with respect to semi-group and structure operations: if $\rho(a_n, b_n) \to 0$, then $\rho(a_n \diamondsuit d, b_n \diamondsuit d) \to 0$ uniformly for every bound set $\mathscr{D} \ni d$, where $\diamondsuit$ is $\odot$ or $\oplus$ or sup or inf.

(3) The minimax condition for the metric

$$\rho(a \diamondsuit b, c \diamondsuit d) \leqslant \max(\rho(a, c), \rho(b, d)),$$

where $\diamondsuit$ stands for $\oplus$ or sup or inf.

(4) The monotonicity of the metric: if $a \geqslant b \geqslant c$, then

$$\rho(a, c) \geqslant \max\{\rho(a, b), \rho(b, c)\}.$$

An important corollary of condition (3) is the minimax inequality

$$\rho\left(\bigoplus_1^n a_i, \bigoplus_1^n b_j\right) \leqslant \min_j \max_i \rho(a_i, b_j). \tag{1.1.5.2}$$

Indeed, induction proves that

$$\rho\left(\bigoplus_1^n a_i, \bigoplus_1^n b_j\right) \leqslant \max\{\rho(a_1, b_{\pi(1)}), \ldots, \rho(a_n, b_{\pi(n)})\}, \tag{1.1.5.3}$$

where π is a permutation of the indices $1, \ldots, n$. Since (1.1.5.2) is valid for every permutation, we select the one for which this maximum is minimal and arrive at (1.1.5.2).

The matching and uniformity conditions are met for the metric $\rho = \rho_{\exp}$. Let us show that the minimax condition is met for $\rho_{\tan^{-1}}$ and for $\rho_{\exp}$. We start with the case $R^1 \cup \{-\infty\}$. We put $a < b$ and $c < d$. Then

$$\rho(a \diamondsuit b, c \diamondsuit d) = \begin{cases} \rho(b, d) & \text{for operations sup and } \oplus. \\ \rho(a, c) & \text{for operation inf.} \end{cases}$$

The multidimensional case can be reduced to the one-dimensional, since for $\rho_{\exp}$ we have

$$\rho(a \diamondsuit b, c \diamondsuit d) \overset{\text{def}}{=} \max_i (\rho((a \diamondsuit b)_i, (c \diamondsuit d)_i)$$
$$= \max_i \rho(a_i \diamondsuit b_i, c_i \diamondsuit d_i).$$

Theorem 1.5.1 *Suppose that f: $\Omega \to A$ is a function with a separable set of values. If f is measurable below, then there exists a sequence of simple functions that converges to f below. If Ω is an A-regular set, A is a locally compact space, metric ρ is minimax, and f is a bounded semi-continuous-below function with a compact support, then there exists a sequence continuous functions $\{f_n\}$ converging to f below at every point.*

Proof. Suppose that $\{f_i^\varepsilon\}_1^\infty$ is a lower ε-mesh on the set of values of f. By $f_i(\omega)$ we denote an elementary function equal to f_i^ε on the set $\Omega_i^\varepsilon = \{\omega: f(\omega) \geqslant f_i^\varepsilon\} \in \mathfrak{A}$ and zero on the complement of Ω_i^ε. Clearly, $f_i(\omega) \leqslant f(\omega)$ and $f_j(\omega) \oplus f(\omega) = f(\omega)$.

Let $f_s^\varepsilon(\omega) = \bigoplus_1^\infty f_i(\omega)$ be a simple function. Then

$$f_s^\varepsilon(\omega) \leqslant f_s^\varepsilon(\omega) \bigoplus_{i=1}^\infty f_i(\omega) \leqslant \bigoplus_1^\infty (f_i(\omega) \oplus f(\omega)) = f(\omega).$$

According to the definition of an ε-mesh, for each point ω there exists a number $i = i(\omega)$ such that

$$\varepsilon \geqslant \rho(f_i^\varepsilon, f(\omega)) = \rho(f_i(\omega), \quad f(\omega)) \geqslant \rho(f_s^\varepsilon(\omega), f(\omega)).$$

Hence,

$$\sup_{\omega \in \Omega} \rho((f_s^\varepsilon(\omega), f(\omega)) < \varepsilon, \quad f_s^\varepsilon(\omega) \leqslant f(\omega).$$

Let us now prove the second part of the theorem. The compactness of the support of a bound function f semi-continuous below implies the relative compactness of the open sets $\Omega_i^\varepsilon = \{\omega, f(\omega) > g_i^\varepsilon\}$, where $\{g_i^\varepsilon\}_1^N$ is the lower ε-mesh on the set of values of f, with $N = N(\varepsilon)$.

Suppose that $\{\mathscr{D}_{i,n}^\varepsilon\}$ is a sequence of compact sets such that $\mathscr{D}_{i,n}^\varepsilon \subset \mathscr{D}_i^\varepsilon$, and $\bigcup\limits_{n=1}^{\infty} \mathscr{D}_{i,n}^\varepsilon = \Omega_i^\varepsilon$. Since it is assumed that Ω constitutes an A-regular space, there are continuous functions $f_{n,i}(\omega) \leqslant g_i^\varepsilon$ that are equal to g_i^ε for $\omega \in \mathscr{D}_{i,n}^\varepsilon$ and to $\mathbb{O}$ for $\omega \notin \mathscr{D}_{i,n}^\varepsilon$. Consider the function $f_n(\omega) = \bigoplus\limits_1^N f_{n,i}^\varepsilon(\omega)\,|_{\varepsilon=n^{-1}}$. Clearly, $f_{n,i}(\omega) \leqslant f(\omega)$ implies $f_n(\omega) \leqslant f(\omega)$. The sum ⊕ of a finite number of functions continuous with respect to the minimax metric is also continuous, whereby $f_n(\omega) \in C(\Omega)$. By virtue of the definition of an ε-mesh, for each point ω there exists a point $q_{i,n}^\varepsilon$ $(i = i(\omega)$ of the ε-mesh such that $f(\omega) \geqslant g_i^\varepsilon$ and $\rho(f(\omega), g_i^\varepsilon) \leqslant \varepsilon$. Thus, $\rho(f(\omega), \bigoplus\limits_1^N f_{n,i}^\varepsilon(\omega)) \leqslant \varepsilon$, with the result that $\rho(f(\omega), f_n(\omega)) \leqslant n^{-1}$. The proof of the theorem is complete.

The quantity

$$\int_\Omega^\oplus f_{\mathrm{el}}(\omega) \odot m(d\omega) = f \odot m(g)$$

is said to be the integral of an A-valued elementary function $f_{\mathrm{el}}(\omega) = \{f, \omega \in g; \mathbb{O}, \omega \notin g\}$ with respect to an A-valued finite idempotent measure m: $\mathfrak{A} \to A$, while the integral of a simple bound function $f_s(\omega) = \bigoplus\limits_1^\infty f_{i,\mathrm{el}}(\omega)$ is the sum

$$\int_\Omega^\oplus f_s(\omega) \odot m(d\omega) = \bigoplus_{i=1}^\infty (f_{i,\mathrm{el}} \odot m(g_i)).$$

Corollary 1.5.1 *If f_s and φ_s are two simple bound functions and $f_s \leqslant \varphi_s$, then*

$$\int_{\Omega}^{\oplus} f_s(\omega) \odot m(d\omega) \leqslant \int_{\Omega}^{\oplus} \varphi_s(\omega) \odot m(d\omega).$$

Lemma 1.5.1 *If a sequence of lower ε-meshes $\{f_i^\varepsilon\}_1^\infty$ is given on the set of values of a simple bound function f_s: $\Omega \to A$, then*

$$\bigoplus_1^\infty f_i \odot m(g_i) = \bigoplus_1^\infty f_i \odot m(G_i) = \lim_{\varepsilon \downarrow 0} \bigoplus_1^\infty f_i^\varepsilon \odot m(G_i^\varepsilon), \qquad (1.1.5.4)$$

where $f_s(\omega) = f_i$ for $\omega \in g_i$, $G_i = \{\omega: f_s(\omega) \geqslant f_i\} = \bigcup_{j:\, f_j \geqslant f_i} g_j$ and $G_i^\varepsilon = \{\omega: f_s(\omega) \geqslant f_i^\varepsilon\}$.

Proof. The set of values of the simple function f_s is discrete, whereby each set G_i^ε is a union of the sets $G_i = \{\omega: f_s(\omega) \geqslant f_i\}$, or $G_i^\varepsilon = \bigcup_{f_j \geqslant f_i^\varepsilon} G_j \supset G_i$. Hence

$$\bigoplus_1^\infty (f_i^\varepsilon \odot m(G_i^\varepsilon)) = \bigoplus_{i=1}^\infty (f_i^\varepsilon \odot (\bigoplus_{j:\, f_i^\varepsilon \leqslant f_j} m(G_j)))$$
$$= \bigoplus_{j=1}^\infty m(G_j) \odot (\bigoplus_{i:\, f_i^\varepsilon \leqslant f_j} f_i^\varepsilon),$$

which, by virtue of the definition of a lower ε-mesh, implies

$$\rho(f_j, \bigoplus_{i:\, f_i^\varepsilon \leqslant f_j} f_i^\varepsilon) \leqslant \varepsilon.$$

The minimax condition for the metric leads to a second inequality:

$$\rho(\bigoplus_1^\infty f_i \odot m(G_i), \bigoplus_1^\infty f_i^\varepsilon \odot m(G_i^\varepsilon))$$
$$\leqslant \max_j \rho(f_j \odot m(G_j), (\bigoplus_{i:\, f_i^\varepsilon \leqslant f_j} f_i^\varepsilon) \odot m(G_i)) \to 0 \quad (1.1.5.5)$$

as $\varepsilon \to 0$, since $m(G_i) \leqslant m(\Omega)$ and measure m is finite. The proof of the lemma is complete.

Finally, we define an integral of a bound measurable-below function f: $\Omega \to A$ with a separable range as the limit of a sequence of integrals of simple functions converging to f below:

$$\int_{\Omega}^{\oplus} f(\omega) \odot m(d\omega) = \lim_{\varepsilon \to 0} \int_{\Omega}^{\oplus} f^\varepsilon(\omega) \odot m(d\omega), \qquad (1.1.5.6)$$

where f^ε constitute the sequence of simple functions set up in the proof of Theorem 1.1.5.1:

$$f^\varepsilon(\omega) = \bigoplus_1^\infty f_i^\varepsilon(\omega),$$

$$f_i^\varepsilon(\omega) = \{f_i^\varepsilon,\ \omega \in \Omega_i^\varepsilon;\ \mathbb{0},\ \omega \notin \Omega_i^\varepsilon\}.$$

$$\Omega_i^\varepsilon = \{\omega\colon\ f^\varepsilon(\omega) \geqslant f_i^\varepsilon\}.$$

The existence of a limit in definition (1.1.5.6) is guaranteed by the following

Theorem 1.1.5.2 *Let A be a partially ordered semi-ring whose metric and structure satisfy conditions* (1)-(4) (see p. 30). *Then for every finite idempotent σ-additive measure and for every bound measurable-below function $f\colon \Omega \to A$ with a separable range the integral* (1.1.5.6) *is finite.*

Proof. Suppose that $\{f_i^\varepsilon\}_1^\infty$ and $\{d_i^\delta\}_1^\infty$ are the lower ε- and δ-meshes on the set of values of map f. Suppose in addition that $f_s^\varepsilon(\omega)$ and $f_s^\delta(\omega)$ are two simple functions corresponding to the ε- and δ-meshes, and $I_\varepsilon = I(f_s^\varepsilon)$ and $I_\delta = I(f_s^\delta)$ are the idempotent integrals of the simple functions. Our aim is to verify that $\rho(I_\varepsilon, I_\delta) \to 0$ as $\varepsilon, \delta \to 0$, that is, $\{I_\varepsilon\}$ is a Cauchy sequence when $\varepsilon \to 0$.

Let us consider the integral $I_{\varepsilon,\delta} = I_\varepsilon \oplus I_\delta = I(f_s^\varepsilon \oplus f_s^\delta)$, which corresponds to the union of the two meshes. Since $\{f_i^\varepsilon\}_{i=1}^\infty$ is the lower ε-mesh, each point d_j^δ belongs to the upper ε-neighborhood of a point f_i^ε. The set of numbers of points of the δ-mesh belonging to the upper ε-neighborhood of point f_i^ε will be denoted by $J_\varepsilon(i)$: $d_j^\varepsilon \geqslant f_i^\varepsilon$, $\rho(f_i^\varepsilon, d_j^\varepsilon) \leqslant \varepsilon\ \forall\ j \in J_\varepsilon(i)$. Interchanging the order of summation, we get

$$I_{\varepsilon,\delta} = \bigoplus_{i=1}^\infty (f_i^\varepsilon \odot m(\mathscr{D}_i^\varepsilon)) \bigoplus_{j\in J_\varepsilon(i)} (d_i^\delta \odot m(\mathscr{D}_j^\delta)),$$

where $\mathscr{D}_i^\varepsilon = \{\omega\colon\ f(\omega) \geqslant f_i^\varepsilon\}$ and $\mathscr{D}_j^\delta = \{\omega\colon\ f(\omega) > d_j^\delta\}$. Since $d_j^\varepsilon \geqslant f_i^\varepsilon$ for all $j \in J_\varepsilon(i)$, we have $\mathscr{D}_j^\varepsilon \subseteq \mathscr{D}_i^\varepsilon$. Hence,

$$I_{\varepsilon,\delta} \leqslant \bigoplus_{i=1}^\infty (f_i^\varepsilon \bigoplus_{j\in J_\varepsilon(i)} d_i^\delta) \odot m(\mathscr{D}_i^\varepsilon) \leqslant \bigoplus_{i=1}^\infty F_i^\varepsilon \odot m(\mathscr{D}_i^\varepsilon),$$

where $F_i^\varepsilon = \sup\{f\colon\ f \geqslant f_i^\varepsilon,\ \rho(f, f_i^\varepsilon) \leqslant \varepsilon\}$.

This result can be used to estimate the distance $\rho(I_\varepsilon, I_{\varepsilon,\delta})$. The minimax property of the metric yields

$$\rho(I_\varepsilon, I_{\varepsilon,\delta}) \leqslant \rho((\bigoplus_{i=1}^\infty (f_i^\varepsilon \odot m(\mathscr{D}_i^\varepsilon)), \bigoplus_{i=1}^\infty (F_i^\varepsilon \odot m(\mathscr{D}_i^\varepsilon)))$$
$$\leqslant \sup_i \{\rho(f_i^\varepsilon \odot m(\mathscr{D}_i^\varepsilon), F_i^\varepsilon \odot m(\mathscr{D}_i^\varepsilon)\}.$$

Bearing in mind that $\rho(f_i^\varepsilon, F_i^\varepsilon) \leqslant \varepsilon$ and $m(\mathscr{D}_i^\varepsilon) \leqslant m(\Omega)$, we conclude that $\rho(I_\varepsilon, I_{\varepsilon,\delta}) \to 0$ as $\varepsilon \to 0$.

Reasoning along the same lines, we can prove that $\rho(I_\delta, I_{\varepsilon,\delta}) \to 0$ as $\delta \to 0$.

If now we employ the triangle property, we finally get $\rho(I_\varepsilon, I_\delta) \leqslant \rho(I_\varepsilon, I_{\varepsilon,\delta}) + \rho(I_{\varepsilon,\delta}, I_\delta) \to 0$ as ε and δ tend to zero. The proof of the theorem is complete.

The following theorem constitutes the main result of the present section.

Theorem 1.1.5.3 *Let m be an arbitrary idempotent σ-additive continuation of finite measure μ from algebra Σ to the σ-algebra $\mathfrak{A}$ and m^* the maximal continuation. Then the integrals with respect to m and m^* of any bounded semi-measurable-below function with a separable range of values coincide.*

Proof. Suppose that $\{f_i^\varepsilon\}_1^\infty$ is the upper ε-mesh on the set of map f and $\{\mathscr{F}_i^\varepsilon\}_1^\infty$, the lower ε-mesh: $\mathscr{F}_i^\varepsilon \leqslant f_i^\varepsilon$, $\rho(f_i^\varepsilon, \mathscr{F}_i^\varepsilon) \leqslant \varepsilon$. By virtue of Theorem 1.5.2, the sequences

$$I_\varepsilon^*(f) = \bigoplus_1^\infty f_i^\varepsilon \odot m^*\{\omega:\ f(\omega) \geqslant f_i^\varepsilon\}$$

and

$$J_\varepsilon^*(f) = \bigoplus_1^\infty \mathscr{F}_i^\varepsilon \odot m^*\{\omega:\ f(\omega) \geqslant \mathscr{F}_i^\varepsilon\}$$

are convergent below to the integral $J^*(f) = \int_\Omega^\oplus f(\omega) \odot m^*(d\omega)$,

while the sequences

$$I_\varepsilon(f) = \bigoplus_1^\infty f_i^\varepsilon \odot m\{\omega:\ f(\omega) \geqslant f_i^\varepsilon\}$$

and

$$J_\varepsilon(f) = \bigoplus_1^\infty \mathscr{F}_i^\varepsilon \odot m\{\omega:\ f(\omega) \geqslant \mathscr{F}_i^\varepsilon\}$$

are convergent below to the integral $I(f) = \int_\Omega^\oplus f(\omega) \odot m(d\omega)$.

Since m^* is bounded and ρ is locally uniform with respect to $\odot$, the convergence of $\rho(f_i^\varepsilon, \mathscr{F}_i^\varepsilon) \to 0$ as $\varepsilon \to 0$ and the minimax property of ρ imply $\rho(\widetilde{J}_\varepsilon^*(f), J_\varepsilon^*(f)) \to 0$ as $\varepsilon \to 0$, with $\widetilde{J}_\varepsilon^*(f) = \bigoplus_1^\infty \mathscr{F}_i^\varepsilon \odot m^*\{\omega:\ f(\omega) \geqslant f_i^\varepsilon\}$.

Combining the definition of the lower ε-mesh and the fact that m and m^* coincide on sets of class G, to which all sets of the type

$\{\omega: f(\omega) > \mathcal{F}_i^\varepsilon\}$ with any functions f that are semi-measurable below belong, we obtain

$$\begin{aligned} m^* \{\omega: f(\omega) \geqslant f_i^\varepsilon\} \leqslant m^* \{\omega: f(\omega) > \mathcal{F}_i^\varepsilon\} = m \{\omega: f(\omega) > \mathcal{F}_i^\varepsilon\} \\ \leqslant m \{\omega: f(\omega) \geqslant \mathcal{F}_i^\varepsilon\} \\ \leqslant m^* \{\omega: f(\omega) \geqslant \mathcal{F}_i^\varepsilon\}. \end{aligned} \tag{1.1.5.7}$$

If we multiply this chain of inequalities by $\mathcal{F}_i^\varepsilon$ and integrate the result with respect to i, we get $\tilde{J}_\varepsilon^* (f) \leqslant J_\varepsilon (f) \leqslant J_\varepsilon^* (f)$. Since $\rho (J_\varepsilon^*, \tilde{J}_\varepsilon^*)$ tends to zero as $\varepsilon \to 0$, we conclude that $\rho (J_\varepsilon (f), J_\varepsilon^* (f)) \to 0$. The proof of the theorem is complete.

1.1.6 Idempotent Measures and Functionals with Values in a Semi-ring

We will prove the theorem of the integral representation of a linear continuous A-valued functional. This theorem can be used to prove the Fubini theorem, whose simple corollaries are Duhamel's integral formulas, which express the solution to the nonhomogeneous evolution equation in terms of integrals of solutions to homogeneous equations.

Let us consider the set $C_0^A (X)$ of continuous A-valued functions with a compact support on the locally compact A-normal topological space X. Here A, the normality of space X, by analogy with the common normality, means that for every closed set K and open set $M \supset K$ there exists a continuous A-valued function equal to $\mathbb{1}$ on K and $\mathbb{0}$ outside M. Suppose that m: $C_0^A (X) \to A$ is a nonnegative linear (with respect to the semi-group operations $\oplus$ and $\odot$) A-valued functional that is continuous in the following sense:

$$\begin{aligned} &\rho (m (\varphi), m (f)) \\ &\leqslant \sup_{x \in \mathcal{K}} \rho (C_{\mathcal{K}} \odot \sup (f (x), \varphi (x)), C_{\mathcal{K}} \odot \inf (f (x) \varphi (x))), \end{aligned}$$

where $\mathcal{K}$ is a compact set containing A — supp f and A — supp φ, and $C_{\mathcal{K}}$ is a constant belonging to A and depending on $\mathcal{K} \subset X$.

In what follows we will repeatedly use the following

Lemma 1.1.6.1 *Let ρ be a minimax metric. Then no finite number of operations $\oplus$,* sup, *and* inf *leads us outside the classes of continuous and semi-continuous-below functions with a compact support.*

Proof. We denote $\oplus$, sup, or inf by $\diamondsuit$. Suppose that f and φ are continuous. If the sequence of $x_n \in X$ converges to an $x \in X$, then

$$\begin{aligned} &\rho (f (x) \diamondsuit \varphi (x), f_n (x) \diamondsuit \varphi_n (x)) \\ &\leqslant \max_n \{\rho (f (x), f (x_n)), \rho (\varphi (x), \varphi (x_n))\} \to 0 \end{aligned}$$

as $n \to \infty$, that is, $f \diamondsuit \varphi$ is a continuous function.

If f and φ are semi-continuous below, there exist sequences of continuous functions $\mathscr{F}_n$ and Φ_n that converge pointwise to f and φ below. From $f \geqslant \mathscr{F}_n$ and $\varphi \geqslant \Phi_n$ it follows that

$$f \lozenge \varphi \geqslant \mathscr{F}_n \lozenge \Phi_n = \Psi_n,$$
$$\rho\,(f(x) \lozenge \varphi(x),$$
$$\Psi_n(x)) \leqslant \max\,\{\rho\,(f(x),\ \mathscr{F}_n(x)),\ \rho\,(\varphi(x),\ \Phi_n(x))\} \to 0$$

as $n \to \infty$. Since, by virtue of the first part of Lemma 1.1.6.1, $\{\Psi_n\}$ constitutes a sequence of continuous functions, $(f \lozenge \varphi)(x)$ is semi-continuous below. The proof of the lemma is complete.

We continue measure m: $C_0^A(X) \to A$ onto the set $\Phi_A(X)$ of functions that are semi-continuous below:

$$m^*(f) = \sup_{\varphi}\,\{m(\varphi),\ \varphi \leqslant f,\ \varphi \in C_0^A(X)\}. \tag{1.1.6.1}$$

Let us first see whether this continuation is monotonic.

Lemma 1.1.6.2 *If* $f_{1,2} \in \Phi_A(X)$ *and* $f_1 \leqslant f_2$, *then* $m^*(f_1) \leqslant m^*(f_2)$.

Proof. It is sufficient to verify that if $\{\varphi_n^1\}_{n=1}^{\infty}$ constitutes a sequence of functions belonging to $C_0^A(X)$ and convergent to f_1 below, there exists a sequence $\{\varphi_n^2\}_1^{\infty}$ convergent to f_2 below such that $\varphi_n^2 \geqslant \varphi_n^1$. By Lemma 1.1.6.1 such a sequence can be obtained from a sequence $\{\tilde{\varphi}_n^2\}_1^{\infty}$ according to the formula $\varphi_n^2(x) = \sup\,\{\varphi_n^1(x),\ \tilde{\varphi}_n^2(x)\}$ or $\varphi_n^2(x) = \varphi_n^1(x) \oplus \tilde{\varphi}_n^2(x)$. The proof of the lemma is complete.

Let us define m^* on a set of arbitrary functions as the upper measure

$$m^*(\Psi) = \inf_{f}\,\{m^*(f),\ f \geqslant \Psi,\ f \in \Phi_A(X)\}. \tag{1.1.6.2}$$

Employing the monotonicity of m^* on $\Phi_A(X)$ and Lemma 1.1.6.1, we can easily prove that the upper measure m^* is monotonic on the set of all functions.

Lemma 1.1.6.3 *If* $\Psi_1 \leqslant \Psi_2$, *then* $m^*(\Psi_1) \leqslant m^*(\Psi_2)$.

Let us say that a partially ordered metric semi-ring A is precompact if on every bounded set $\mathscr{L} \subset A$ there exists a finite ε-mesh $\{d_i^{\varepsilon}\}_1^N$, $N = N(\varepsilon)$, that is, such that $\forall x \in \mathscr{L}\ \ \exists d_i^{\varepsilon}$: $\rho\,(x,\ d_i^{\varepsilon}) \leqslant \varepsilon$.

Corollary *If a locally compact semi-ring A is precompact, for every bounded function Ψ there exists a function f semi-continuous below such that $\rho\,(m^*(\Psi),\ m^*(f)) \leqslant \varepsilon$, $f \geqslant \Psi$.*

Indeed, let $\mathscr{L} = \{m^*(f)\colon f \geqslant \Psi,\ f \leqslant \sup \Psi,\ f \in \Phi_A(X)\}$ be a bounded set and $\{d_i^{\varepsilon}\}_1^N$ an ε-mesh on $\mathscr{L}$. Let us also assume that $f_i^{\varepsilon}(x)$ are semi-continuous-below functions belonging to $\Phi_A(X)$ and such that $m(f_i^{\varepsilon}) = d_i^{\varepsilon}$. In this case the function $f^{\varepsilon}(x) =$

$\inf\limits_{1\leqslant i\leqslant N} f_i^\varepsilon(x)$ is semi-continuous below, too (see Lemma 1.1.6.1), $f^\varepsilon \geqslant \Psi$, since $f_i^\varepsilon \geqslant \Psi$ and $\rho(m^*(f^\varepsilon), m^*(\Psi)) \leqslant \varepsilon$ by the definition of an ε-mesh.

The commutativity of the passage to the limit with a linear continuous functional is ensured by the following

Theorem 1.1.6.1 *If a monotone increasing sequence of functions Ψ_n converges below to a bounded function Ψ, then $m^*(\Psi_n) \uparrow m^*(\Psi)$:*

$$\lim\sup m^*(\Psi_n) = m^*(\lim\sup \Psi_n).$$

Proof. Clearly,

$$m^*(\Psi) \geqslant \sup_n m^*(\Psi_n). \tag{1.1.6.3}$$

By virtue of the corollary to Lemma 1.1.6.3, for every positive ε there exists a function $f_n \in \Phi_A(X)$ semi-continuous below such that $\Psi_n \leqslant f_n$, $m^*(\Psi_n) \leqslant m^*(f_n)$, and $\rho(m^*(\Psi_n), m^*(f_n)) \leqslant \varepsilon \times 2^{-n}$. Suppose that $\varphi_n(x) = \sup(f_1(x), \ldots, f_n(x)) \in \Phi_A(X)$. Clearly, $\varphi_{n+1} \geqslant f_{n+1} \geqslant \Psi_{n+1}$, $\varphi_{n+1} \geqslant \varphi_n$. Since $\sup\{\varphi_n, f_{n+1}\} \leqslant \varphi_n \oplus f_{n+1}$ and $\rho(a, b \oplus c) \leqslant \rho(a, b) \oplus \rho(a, c)$, we have

$$\begin{aligned}
&\rho(m^*(\Psi_{n+1}), m^*(\varphi_{n+1})) \leqslant \rho(m^*(\Psi_{n+1}), m^*(\varphi_n)) \\
&+ \rho(m^*(\Psi_{n+1}), m^*(f_{n+1})) \\
&\leqslant \rho(m^*(\Psi_n), m^*(\varphi_n)) + 2^{-n}\varepsilon \leqslant \ldots \\
&\leqslant \rho(m^*(\Psi_1), m^*(\varphi_1)) + (2^{-1} + \ldots + 2^{-n})\varepsilon \leqslant 2\varepsilon.
\end{aligned}$$

Hence, $\sup\limits_n \varphi_n \geqslant \Psi_n$ and

$$\rho(m^*(\Psi_n), m^*(\varphi_n)) \geqslant \rho(m^*(f_n), m^*(\varphi_n)) \leqslant \varepsilon. \tag{1.1.6.4}$$

In view of the concordance of metric and structure and the arbitrariness of the choice of ε, inequality (1.1.6.4) implies

$$\sup_n m^*(\Psi_n) \leqslant m^*(\sup_n \Psi_n) = m^*(\Psi).$$

Indeed, if we assume that the opposite is true, that is, if $m^*(\varphi) > \sup\limits_n m^*(\Psi_n)$, the more so $\sup m^*(\varphi_n) > m^*(\Psi_n)$, and inequality (1.1.6.4) cannot be satisfied for an arbitrary (positive) ε. If we allow for (1.1.6.3), the proof of the theorem is complete.

Lemma 1.1.6.4 (Fatou's lemma) *Let $\{f_n\}$ be a uniformly bounded sequence of A-valued functions. Then*

$$m^*(\lim\inf f_n) \leqslant \lim\inf m^*(f_n).$$

Proof. Let us consider the functions $\mathscr{F}_n(x) = \inf\{f_m(x), m \geqslant n\} \leqslant f_n$. These constitute a nondecreasing sequence that converges,

by definition, to $\liminf f_n$. By Theorem 1.1.6.1,

$$\limsup m^* (\mathcal{F}_n) = m^* (\limsup \mathcal{F}_n) = m^* (\liminf f_n). \tag{1.1.6.5}$$

Since $\mathcal{F}_n (x) \leqslant f_m (x)$ for $m \geqslant n$, we have $m^* (\mathcal{F}_n (x)) \leqslant m^* (f_m)$, $m \geqslant n$, whereby $m^* (\mathcal{F}_n (x)) \leqslant \inf\limits_m m^* (f_m)$. Combining this with (1.1.6.5) yields

$$\liminf m^* (f_n) \geqslant m^* (\liminf f_n).$$

The proof of the lemma is complete.

A set $a \in X$ is said to be integrable if the characteristic function χ_a of a is integrable (i.e. $\rho (\mathbb{O}, m^* (\chi_a)) < \infty$) and measurable if the characteristic functions of $a \cap K$ are integrable for every compact metric space K. A set a is said to be locally negligible if

$$m^* (a) = \sup_k m^* (a \cap K) = \mathbb{O},$$

where $m^* (a) = m^* (\chi_a)$, with X a locally compact topological space.

Theorem 1.1.6.5 implies that if sets a_n are integrable and $m^* (a_n) \leqslant C$, with C a constant, the set $a = \bigcup\limits_1^\infty a_n$ is also integrable, $\chi_a (x) = \bigoplus\limits_1^\infty \chi_{a_i} (x)$, and

$$m^* (\chi_a) = m^* (a) = \limsup m^* \left(\bigoplus_1^n \chi_{a_i} \right)$$
$$= m^* \left(\limsup \bigoplus_1^n \chi_{a_i} \right).$$

Hence, m^* is a σ-additive idempotent measure.

By $\mathcal{B} (X)$ we denote the σ-algebra of Borel sets, that is, the smallest σ-algebra with respect to which all continuous functions are measurable.

Theorem 1.1.6.2 *If φ is a bounded function $\mathcal{B}$-measurable below with a separable range of values, then $m^* (\varphi) = \int\limits_X^{\oplus} \varphi (x) \odot m^* (dx)$, where $m^* (dx)$ is the upper measure on $\mathcal{B} (X)$ generated by the linear continuous functional m on $C_0^A (X)$, with $m^* (\varphi)$ the upper continuation of m.*

Proof. Suppose that $\varphi_n^\varepsilon (x) = \bigoplus\limits_1^\infty \varphi_i^\varepsilon \odot \chi \{x\colon \varphi (x) \geqslant \varphi_i^\varepsilon\}$, where $\{\varphi_i^\varepsilon\}_1^\infty$ is the ε-mesh on the set of values of map φ. Then $\varphi_n^\varepsilon \leqslant \varphi$ and $\rho (\varphi_n^\varepsilon, \varphi) \leqslant \varepsilon$. Thus, $\varphi_n^\varepsilon \uparrow \varphi$. By virtue of Theorem 1.1.6.1,

$m^*(\varphi_n^\varepsilon) \uparrow m^*(\varphi)$, while by virtue of Theorems 1.1.5.1 and 1.1.5.2,

$$\int_X^\oplus \varphi_n^\varepsilon(x) \odot m^*(dx) \uparrow \int_X^\oplus \varphi(x) \odot m^*(dx).$$

Since for simple functions

$$m^*(f^\varepsilon) = \bigoplus_1^\infty f_i \odot m^*(\chi_{a_i}) = \bigoplus_1^\infty f_i \odot m^*(a_i)$$
$$= \int_X^\oplus f^\varepsilon(x) \odot m^*(dx),$$

a similar equation holds for the limit as $\varepsilon \to 0$. The proof of the theorem is complete.

Let m_1 and m_2 be two linear continuous A-valued functionals on $C_0^A(X_1)$ and $C_0^A(X_2)$, respectively, with X_1 and X_2 two locally compact A-regular topological spaces, let $X = X_1 \times X_2$ be the direct product of these locally compact spaces, and let $C_0^A(X)$ be the set of continuous A-valued functions on $X = X_1 \times X_2$ with a compact support. It is easy to see that the functions $\varphi_1(x_1) = m_2(\varphi(x_1, \cdot))$ and $\varphi_2(x_2) = m_1(\varphi(\cdot, x_2))$ belong to $C_0^A(X_1)$ and $C_0^A(X_2)$, respectively, if $\varphi \in C_0^A(X)$. Hence, the expressions $m_{12}(\varphi) = m_1(m_2(\varphi))$ and $m_{21}(\varphi) = m_2(m_1(\varphi))$ define linear continuous A-valued functionals on $C_0^A(X)$.

Our immediate task is to prove that $m_{12} = m_{21}$.

Theorem 1.1.6.3 *Let $f \in C_0^A(X)$, $X = X_1 \times X_2$, be a continuous function with a separable range of values. Then $m_{12}(f) = m_{21}(f)$.*

Proof. Suppose that K_1 and K_2 are two compact subsets in X_1 and X_2 whose product contains A — supp f. The function $\rho(f(x_1, x_2), f(x_1', x_2'))$ is continuous on the compact metric space $K = (K_1 \times K_1) \times (K_2 \times K_2)$, whereby it is equicontinuous, and for every positive ε there exist finite nonintersecting partitions of compact metric spaces K_1 and K_2 such that

$$\rho(f(x_1, x_2), f(x_1', x_2')) \leqslant \varepsilon \quad \forall x_1 \in K_1,\ x_2,\ x_2' \in K_{2,j}^\varepsilon,$$
$$\rho(f(x_1, x_2), f(x_1', x_2')) \leqslant \varepsilon \quad \forall x_2 \in K_2,\ x_1,\ x_1' \in K_{1,j}^\varepsilon,$$
$$K_1 = \bigcup_{j=1}^N K_{1,j}^\varepsilon, \quad K_2 = \bigcup_{j=1}^N K_{2,j}^\varepsilon, \quad N = N(\varepsilon).$$

Let us consider the sequence of simple functions $f_\varepsilon(x_1, x_2) = \bigoplus_{i,j} \varphi_{ij}^\varepsilon(x_1, x_2)$, with

$$\varphi_{ij}^\varepsilon(x_1, x_2) = \begin{cases} \inf\limits_{K_{1,i}^\varepsilon \times K_{2,j}^\varepsilon} & \text{if } (x_1, x_2) = \varphi_{ij}^\varepsilon \text{ if } (x_1, x_2) \in K_{1,i}^\varepsilon \times K_{2,j}^\varepsilon, \\ \mathbb{O} & \text{if } (x_1, x_2) \notin K_{1,i}^\varepsilon \times K_{2,j}^\varepsilon, \end{cases}$$

which converges uniformly to $f(x_1, x_2)$ below. By virtue of Theorem 1.1.6.2, the sequences of the simple functions

$$\Psi^\varepsilon(x_1) = \bigoplus_{i=1}^{N} \Psi_i^\varepsilon(x_1),$$

$$\Psi_i^\varepsilon(x_1) = \begin{cases} \bigoplus_{j=1}^{N} \varphi_{ij}^\varepsilon \odot m_2(K_{2,j}^\varepsilon) & \text{if} \quad x_1 \in K_{1,i}^\varepsilon, \\ \mathbf{0} & \text{if} \quad x_1 \notin K_{1,i}^\varepsilon, \end{cases}$$

$$\Phi^\varepsilon(x_2) = \bigoplus_{j=1}^{N} \Phi_j^\varepsilon(x_2),$$

$$\Phi_j^\varepsilon(x_2) = \begin{cases} \bigoplus_{i=1}^{N} \varphi_{ij}^\varepsilon \odot m_1(K_{1,i}^\varepsilon) & \text{if} \quad x_2 \in K_{2,j}^\varepsilon, \\ \mathbf{0} & \text{if} \quad x_2 \in K_{2,j}^\varepsilon, \end{cases}$$

converge, as $\varepsilon \to 0$, to the integrals

$$\Psi(x_1) = \int_{X_2}^{\oplus} f(x_1, x_2) \odot m_2(dx_2) = m_2(f(x_1, \cdot)),$$

$$\Phi(x_2) = \int_{X_1}^{\oplus} f(x_1, x_2) \odot m_1(dx_1) = m_1(f(\cdot, x_2)).$$

But for every finite $N = N(\varepsilon)$ the integrals $\int_{X_1}^{\oplus} \Psi^\varepsilon(x_1) \odot m_1(dx_1)$ and $\int_{X_2}^{\oplus} \Phi^\varepsilon(x_2) \odot m_2(dx_2)$ are equal to the same sum, $\bigoplus_{i,j=1}^{N} \varphi_{ij}^\varepsilon \odot m_1(K_{1,i}^\varepsilon) \odot m_2(K_{2,j}^\varepsilon)$. Hence, a similar equality is valid in the limit, too. The proof of the theorem is complete.

The result obtained for continuous functions can be extended so as to incorporate the cases of a set of bounded semi-continuous-below functions with a separable range of values and a set of all bounded measurable functions.

1.1.7 The Fourier-Legendre Transform

We will show that with respect to the semi-group operations of addition and multiplication the Laplace transform has the same meaning as the Fourier transform with respect to the arithmetical operations of addition and multiplication.

Let the locally convex A-regular topological space X be an Abelian group and let m be the linear continuous A-valued functional $C_0^A(X)$

that is invariant with respect to the group of translations, or $m(T_q\varphi) = m(\varphi)$, where $T_q\varphi(x) = \varphi(x+q)$, $x, q \in X$. By Theorem 1.1.6.3, the functional m admits, on the set of functions on $C_0^A(X)$ with a separable set of values, the integral representation

$$m(\varphi) = \int_X^{\oplus} \varphi(x) \odot m^*(dx),$$

where m^* is the upper measure on $\mathscr{B}(X)$ generated by m.

Let φ and Ψ be two functions belonging to $C_0^A(X)$. The convolution of these two functions is a function from $C_0^A(X)$ defined thus:

$$(\varphi * \Psi_m)(q) = m(T_q\varphi' \odot \Psi), \tag{1.1.7.1}$$

where $\varphi'(x) = \varphi(-x)$, and $T_q\varphi'(x) = \varphi(q-x)$. Clearly,

$$T_y(\varphi * \Psi)(q) = (T_{-y}\varphi * \Psi)(q) = (T_{-y}\Psi * \varphi)(q). \tag{1.1.7.2}$$

Let us consider the linear continuous operator $\mathscr{F}_{x\to p}$ that maps set $C_0^A(X)$ into the set of continuous bounded A-valued functions, $C^A(X)$, and possesses the following characteristic property: operator $\mathscr{F}_{x\to p}$ is the equioperator of the translation operator, or $\mathscr{F}_{x\to p}T_q = e(p, q) \odot \mathscr{F}_{x\to p}$, where $e(p, q)$ is a continuous bounded A-valued function. The Fourier transform of functions in R^n has just this property, with $e(p, q) = \exp\{\pm i(p, q)\}$.

In general,

$$\begin{aligned}\mathscr{F}_{x\to p}(T_{q+u}\varphi) &= e(q+u, p) \odot \mathscr{F}_{x\to p}\varphi\\ &= \mathscr{F}_{x\to p}(T_q \circ T_u\varphi)\\ &= e(p, q) \odot e(p, u) \odot \mathscr{F}_{x\to p}\varphi,\end{aligned}$$

from which it follows that

$$e(p, q) \odot e(p, u) = e(p, q+u). \tag{1.1.7.3}$$

Another important property of the common Fourier transform is the formula for commutation with the convolution:

$$\mathscr{F}_{x\to p}(\varphi * \Psi) = (\mathscr{F}_{x\to p}\varphi) \odot (\mathscr{F}_{x\to p}\Psi). \tag{1.1.7.4}$$

Let us show that if formula (1.1.7.4) is taken as a condition, then the function e satisfies condition (1.1.7.3) in the second independent variable, too, and the Fourier transform admits an integral representation of operator $\mathscr{F}$ on a set of functions with a separable range

of values. Using the convolution property (1.1.7.1), we get

$$\begin{aligned}(\mathcal{F}_{x\to p}\varphi) \odot (\mathcal{F}_{x\to p}\Psi) &= \mathcal{F}_{x\to p}(\varphi * \Psi)\\ &= m(\mathcal{F}_{x\to p}(T_-(\cdot)\varphi) \odot \Psi(\cdot))\\ &= (\mathcal{F}_{x\to p}\varphi) \odot m(e(-p, \cdot) \odot \Psi(\cdot)),\end{aligned}$$

where $(\cdot)$ stands for the independent variable in which m operates. Hence, $\mathcal{F}_{x\to p}\Psi = m(e(-p, \cdot) \odot \Psi(\cdot))$. For continuous functions with a separable set of values, we can write this relationship in the form of an integral:

$$\mathcal{F}_{x\to p}\Psi = \int_X^{\oplus} \Psi(x) \odot e(-p, x) \odot m^*(dx), \qquad (1.1.7.5)$$

where $(e(-p-p', x+x') = e(-p, x) \odot e(-p', x) \odot$ $e(-p, x') \odot e(-p', x')$.

Example. Suppose that $A = R \cup \{-\infty\}$, $X = R^n$, $\oplus = \sup$, $\odot = +$, $m = \text{const}$, $\mathbb{1} = 0$, and $\mathbb{0} = -\infty$. Then $\mathcal{F}_{x\to p}\varphi = \sup\limits_X (e(p, x) + \varphi(x))$, where $e(x, p)$ is a continuous additive function of independent variables x and p. Every continuous additive function is linear, whereby $e(p, x) = (p, Hx)$, with H an n-by-n matrix. Thus, here the Fourier transform in the sense of semi-ring A coincides with the Legendre transform in the sense of group operations in R.

1.1.8 Duhamel's Theorem

As another application of the theory of idempotent integration we consider the proof of Duhamel's theorem in the semi-group case, whichexcludes subtraction and differentiation. To avoid differentiation one can go over to the integral representation and employ Fubini's theorem, which justifies the change in the order of integration.

Suppose that we have specified a bounded continuous-in-t family of linear operators $\hat{L}(t)$, $t \in R$, acting in a normed space of A-valued functions, $\mathcal{B}(X)$, that is closed with respect to the operations of addition, $\oplus$, and multiplication by a constant, $\odot$. Suppose that m: $\mathcal{B}(R) \to A$ is an A-valued idempotent σ-additive bounded measure and $\varphi_t(x)$, $\Psi_{t,s}(x)$, and $\Phi_t(x)$ are A-valued functions that belong to $\mathcal{B}(X)$ are continuous in parameter $t \in R$, and have separable ranges of values. If the operators $\hat{L}(t)$ retain the separability of the ranges of values (say, if the entire space $\mathcal{B}(X)$ is separable), then the integrals $\int_{(0,t]}^{\oplus} \hat{L}(\tau)\varphi_\tau \odot m(d\tau)$ and $\int_{(0,t]}^{\oplus} \hat{L}(\tau)\Psi_{\tau,s} \odot m(d\tau)$ have finite values. Let us assume that the functions φ_t, $\Psi_{t,s}$, and Φ_t are solutions to the equations listed in Table 1.1.1.

Table 1.1.1

Integral representation of equations for semi-group operations $\oplus$ and $\odot$	Differential representation of equations for group operations $+$ and $\times$	
$\varphi_t = \varphi_0 \oplus \int\limits_{(0,t]}^{\oplus} (\hat{L}(\tau)\varphi_\tau \oplus \psi_\tau) \odot m(d\tau)$	$\frac{\partial \varphi_t}{\partial t} = \hat{L}(t)\varphi_t + \Psi_t,$ $\varphi\mid_{t=0} = \varphi_0 \in \mathscr{B}(X)$	(1.1.8.1)
$\Psi_{t,s} = \Psi_s \oplus \int\limits_{(s,t]}^{\oplus} (\hat{L}(\tau)\Psi_{\tau,s}) \odot m(d\tau)$	$\frac{\partial \Psi_{t,s}}{\partial t} = \hat{L}(t)\Psi_{t,s},$ $\Psi\mid_{t=s} = \Psi_s \in \mathscr{B}(X)$	(1.1.8.2)
$\Phi_t = \varphi_0 \oplus \int\limits_{(0,t]}^{\oplus} (\hat{L}(\tau)\Phi_\tau) \odot m(d\tau)$	$\frac{\partial \Phi_t}{\partial t} = \hat{L}(t)\Phi_t,$ $\Phi\mid_{t=0} = \varphi_0.$	(1.1.8.3)

Theorem 1.1.8.1 *Let two functions,* Ψ *and* Φ, *satisfy the integral equations* (1.1.8.2) *and* (1.1.8.3). *Then the function*

$$S_t = \Phi_t \oplus \int\limits_{(0,t]}^{\oplus} \Psi_{t,\tau} \odot m(d\tau) \tag{1.1.8.4}$$

is a solution to Eq. (1.1.8.1).

Proof. Using the integral representations for Φ_t and $\Psi_{t,\tau}$, we arrive at the following identity:

$$S_t = \varphi_0 \oplus \int\limits_{(0,t]}^{\oplus} (\hat{L}(\tau)\Phi_\tau) \odot m(d\tau)$$
$$\oplus \int\limits_{(0,t]}^{\oplus} \Big(\Psi_\tau \oplus \int\limits_{(\tau,t]}^{\oplus} (\hat{L}(s)\Psi_{s,\tau}) \odot m(ds)\Big) \odot m(d\tau). \tag{1.1.8.5}$$

Note that the function $f(s, \tau) = \hat{L}(s)\Psi_{s,\tau}$ and the measure $\mu = m \times m$ on $R \times R$ satisfy the hypothesis of Fubini's theorem, whereby the values of the integral over the domain $\Omega = \{(s, \tau): s \geqslant \tau\}$ does not depend on the order of integration:

$$\int\limits_{(0,t]} m(d\tau) \odot \int\limits_{(\tau,t]} (\hat{L}(s)\Psi_{s,\tau}) \odot m(ds)$$
$$= \int\limits_{(0,t]} m(ds) \odot \int\limits_{(0,s]} \hat{L}(s)\Psi_{s,\tau} \odot m(d\tau).$$

Combining this with (1.1.8.5) yields

$$S_t = \varphi_0 \oplus \int\limits_{(0,\, t]}^{\oplus} \left\{ \hat{L}(\tau) \left(\Phi_\tau \oplus \int\limits_{(0,\, \tau]}^{\oplus} \Psi_{\tau,\, s} \odot m(ds) \oplus \Psi_\tau \right) \right\} \odot m(d\tau)$$

$$= \varphi_0 \oplus \int\limits_{(0,\, t]}^{\oplus} (\hat{L}(\tau) S_\tau \oplus \Psi_\tau) \odot m(d\tau).$$

The proof of the theorem is complete.

1.1.9 The Fredholm Alternative

In this section we show how a generalized solution can be determined by introducing a separable scalar product and constructing a conjugate operator. The condition for the uniqueness of the solution to a nonhomogeneous equation will be formulated as the Fredholm alternative.

Let $\mathcal{L}$ be the space of bounded-above functions v: $X \to R$ that determine the idempotent measures $\mu(b) = \sup\limits_{x \in b} v(x)$ on the σ-algebra $\mathscr{B}(X)$. Let M be the set of bounded semi-measurable-below functions f: $X \to R \cup \{-\infty\}$ for which, by Theorem 1.1.5.1, the idempotent Lebesgue integral $(f, v) = \int\limits_X^{\oplus} f(x) \odot \mu(dx)$ has a finite value. Finally, on $\mathcal{L}$ we define the equivalence relation $v_1 \sim v_2 \Leftrightarrow (f, v_1) = (f, v_2) \;\; \forall f \in M$.

Theorem 1.1.9.1 *The equivalence relation defined on $\mathcal{L}$ is the congruence of A, a semi-module of space $\mathcal{L}$.*

Proof. Indeed, if $v_1 \sim v_2$, then $a \odot v_1 \sim a \odot v_2$ for every $a \in A$, since $(f, a \odot v_1) = (a \odot f, v_1) = (a \odot f, v_2) = (f, a \odot v_2)$. Reasoning along the same lines, we can also verify that if $v_1 \sim v_1'$ and $v_2 \sim v_2'$, then $v_1 \oplus v_2 \sim v_1' \oplus v_2'$.

Note that this theorem holds true for every semi-module of A-valued functions $v \in \mathcal{L}$ and for the respective dual semi-module of the homomorphisms $f \in M$ in the semi-ring A with respect to the scalar product $(f, v) = f(v)$, with every congruence on $\mathcal{L}$ being determined by a dual semi-module M, for which we may take the maximal semi-module formed by all homomorphisms f: $\mathcal{L} \to A$ for which $v_1 \sim v_2 \Rightarrow f(v_1) = f(v_2)$.

Let us consider the equation $Tu = v$ in $\mathcal{L}$. With v running through $\mathcal{L}$, the set of the solutions to this equation form a quotient semi-module with respect to the congruence $u_1 \sim u_2 \Leftrightarrow Tu_1 = Tu_2$ corresponding to the dual semi-module $M = T^*\mathcal{L}^*$, with T^* the conjugate operator. Indeed, if $Tu_1 = v = Tu_2$, then $(T^*g)(u_1) = g(Tu_1) = g(Tu_2) = (T^*g)(u_2)$ for every $g \in \mathcal{L}^*$, that is, $f(u_1) = f(u_2)$ for every $f \in T^*\mathcal{L}^*$, and vice versa, if $f(u_1) = f(u_2)$ for every $f \in T^*\mathcal{L}^*$, then $Tu_1 = Tu_2$, that is, u_1 and u_2 are solutions to

equations with the same right-hand side. We have therefore proven the following

Theorem 1.1.9.2 *Let $M = T^*\mathscr{L}^*$ be the range of values of the conjugate operator T^*, that is, the equation $T^*g = f$ has a solution for every $f \in M$. Then the solution to the equation $Tu = v$ is unique to within the congruence determined by semi-module M.*

The reverse statement is also true if an additional condition is imposed on space M. We will say that a dual semi-module is completely separating with respect to congruence $\sim$ if for every proper subspace $M_1 \subset M$ there exist elements $u_1, u_2 \in \mathscr{L}$ that are not equivalent to mod M and for which $f(u_1) = f(u_2)\ \forall f \in M_1$.

Theorem 1.1.9.3 *Let the equation $Tu = v$ have a unique solution to within an equivalence relation determined by a completely separating dual semi-module M. Then $T^*\mathscr{L}^* = M$, that is, the equation $T^*g = f$ has a solution for every $f \in M$.*

Proof. Let $Tu_1 = v = Tu_2$, that is,

$$T^*g(u_1) = gT(u_1) = gT(u_2) = T^*g(u_2)\ \forall g \in \mathscr{L}^*.$$

Then, according to the hypothesis, $u_1 \sim u_2 \pmod M$, that is, $T^*g \in M$ for every $g \in \mathscr{L}^*$, whereby $T^*\mathscr{L}^* \subseteq M$. Let us prove that $T^*\mathscr{L}^* = M$. Suppose that the opposite is true, that is, suppose that $T^*\mathscr{L}^* \subset M$. Then, in view of the condition that M be completely separating, there exist $u_1 \not\sim u_2$ such that $(T^*g, u_1) = (T^*g, u_2)$ for every $g \in \mathscr{L}^*$. But this implies that $Tu_1 = Tu_2$ in view of the arbitrariness of $g \in \mathscr{L}^*$, which means that the uniqueness condition imposed on the solution of the equation $Tu = v$ is violated.

1.1.10 The Generalized Discrete Bellman Equation

Here we introduce the concept of the generalized Bellman equation in a discrete medium.

Let $X = \{1, \ldots, n\}$ be a finite set equipped with a discrete topology. Then continuous functions $\varphi\colon X \to A$ can be identified with vectors $a = (a_1, \ldots, a_n) \in A^X$ that have arbitrary components $a_i \in A$. The general form of the functional (measure) and the operator (endomorphism) m that maps functions with values in semi-ring A^x into functions with values with semi-sing A is established through the following

Theorem 1.10.1 *Every homomorphism $m\colon A^X \to A$ (functional) has the form $m(a) = m^1 \odot a_1 \oplus \ldots \oplus m^n \odot a_n$, where $m^i \in A$ are arbitrary elements of semi-ring A. The positive measure m is determined by the positive elements $m^i \geqslant \mathbb{0}$.*

Proof. If we write vectors a in the form $a = a_1 \odot e^1 \oplus \ldots \oplus a_n \odot e^n$, where e^i is a row in which all components are zeros except the

ith component, which is equal to unity, then, by virtue of linearity,

$$\mu(a) = a_1 \odot \mu(e^1) \oplus \ldots \oplus a_n \odot \mu(e^n).$$

If we introduce the notation $\mu(e^i) = m^i$, we arrive at the sought representation. Conversely, if $m^i \in A$ are elements of semi-ring A, the combination $\mu(a)$ clearly defines a measure on $C_0(X) = A^X$.

Corollary *Every endomorphism G: $A^X \to A^X$ of semi-module A^X has the form*

$$G(a)_i = g_i^1 \odot a_1 \oplus \ldots \oplus g_i^n \odot a_n, \quad 1 \leqslant i \leqslant n,$$

where the g_i^j are arbitrary elements of A.

Such a "source-wise representation" is obtained by applying the theorem on the representation of measure μ on A^X for each component $G(a)_i$.

What will be called the generalized discrete Bellman equation is the evolution equation in discrete time,

$$L_t\colon C_0^A(X) \to C_0^A(X), \quad \varphi_{t+1} = L_t\varphi_t \tag{1.1.10.1}$$

defined by operators of the type L_t: $C_0^A(X) \to C_0^X(X)$. Operator L_t can be written, via the generalized matrix elements $\lambda_t^{x_0}(x)$, in integral form thus:

$$\varphi_{t+1}(x_0) + \int^{\oplus} \lambda_t^{x_0}(x) \odot \varphi_t(x) \odot m(dx),$$

or, in the case of a finite set x, in matrix form:

$$a_{j,t+1} = e_{j,t}^1 \odot a_{1,t} \oplus \ldots \oplus e_{j,t}^n \odot a_{n,t}.$$

The solution to the Bellman equation is given by the formula $\varphi_t = G_t(t_0)\,\varphi_{t_0}$, where operator $G_t(t_0) = L_{t-1} \circ \ldots \circ L_{t_0+1} \circ L_{t_0}$ constitutes a composition of operators L_t.

Example 1. Let us consider the case of $A = R \cup \{-\infty\}$, $a \oplus b = \max(a, b)$, $a \odot b = a + b$, a zero $\mathbb{0} = -\infty$, and an identity $\mathbb{1} = 0$. Suppose that $m(\varphi) = \sup \varphi(x) = \max \varphi(x)$ is a positive measure on $C_0(X)$ that determines the scalar product $(\varphi, \Psi) = \max(\varphi(x) + \Psi(x))$. Obviously, the linearity condition imposed on the functional $\varphi \to \mu(\varphi)$ is met:

$$\begin{aligned} m(\max\{a + \varphi, b + \Psi\}) &= \sup_x \max\{a + \varphi, b + \Psi\} \\ &= \max(a + \sup_x \varphi(x), b + \sup_x \Psi(x)). \end{aligned}$$

From this follows the linearity (with respect to the semi-ring considered here) of the generating operator L for the common Bellman

equation in discrete time,

$$\varphi_{t+1}(x) = \sup_y (\varphi_t(y) + \lambda_t^x(y)),$$

which is determined by the family $\{e_t^x\}$ of linear functionals

$$e_t^x(\varphi) = \sup_y (\varphi(y) + \lambda_t^x(y)).$$

Example 2. Consider the semi-ring $A = R \cup \{\pm\infty\}$, $a \oplus b = \max(a, b)$, $a \odot b = \min(a, b)$, with a zero $\mathbb{O} = -\infty$ and an identity $\mathbb{1} = \infty$. Let us assume that $m(\varphi) = \sup \varphi(x) = \max \varphi(x)$ is the positive measure on $C_0(X)$, whose linearity can be verified directly:

$$\begin{aligned} m(\max\{\min(a, \varphi), \min(b, \Psi)\}) &= \sup_x \max(\min(a, \varphi(x)), \min(b, \Psi(x))) \\ &= \max(\min(a, \sup_x \varphi(x)), \min(b, \sup_x \Psi(x))) \\ &= \max(\min(a, m(\varphi)), \min(b, m(\Psi))). \end{aligned}$$

We now write the scalar product using the measure m thus:

$$(\varphi, \Psi) = \sup_X \min(\varphi(x), \Psi(x)).$$

Substituting $\Psi(x) = \lambda_t^y(x)$, where $\lambda_t^y(x)$ is the kernel of the evolution equation

$$\varphi_{t+1}(y) = \sup_X \min(\varphi_t(x), \lambda_t^y(x)),$$

we conclude that the "minimax" Bellman equation is also linear in the space considered here.

The evolution equation (1.1.10.1) for the semi-rings considered in Examples 1 and 2 is commonly interpreted in optimization problems of discrete mathematics as the equation for describing the evolution in a discrete medium. Let us now define the appropriate concepts for the case of an arbitrary semi-ring.

We denote the elements of a finite set X by $x_1, x_2, \ldots, x_N$, $|x| = N$, and call them the points of the medium. An ordered pair of points of the medium (x_i, x_j) for which $L(i, j) = L(x_i, x_j) \neq \mathbb{O}$ we will call the connection of points x_i and x_j. Let us denote the set of all connections of the points of the medium, $\Gamma = \{(x_i, x_j), L(x_i, x_j) \neq \mathbb{O}\}$ by $\Gamma \subseteq X \times X$. We will call map $L_M\colon \Gamma \to A \setminus \{\mathbb{O}\}$ the characteristic of connections or the characteristic of the functions of the medium. The collection of objects $(X, \Gamma, L, A) = \mathcal{M}$ will be called a discrete medium.

Let us now define a neighbourhood $\Gamma(x_i)$ of a point x_i as the set of points of the medium $\mathcal{M}$ with which x_i is connected, that is, $\Gamma(x_i) = \{y \in X\colon (x_i, y) \in \Gamma\}$. The space of states H_M of medium $\mathcal{M}$

is defined as the semi-module $C(X) = \{\varphi: X \to A\}$. The one-parameter family of states $S = \{S_t, t = 0, 1, 2, \ldots\}$ is called an evolution in discrete time in medium $\mathcal{M}$ (S_0 is the initial state in the evolution) if states that are one step distant in time t are related thus:

$$S_{t+1} = \hat{L}_M S_t, \quad i = 0, 1, \ldots, \tag{1.1.10.2}$$

where the endomorphism $\hat{L}_M$: $H_M \to H_M$ is determined by the characteristic function of the medium through the formula

$$(\hat{L}_M \varphi)(x_i) = \bigoplus_{x'_j \in \Gamma(x_i)} L(x_i, x'_j) \odot \varphi(x'_j). \tag{1.1.10.3}$$

Equation (1.1.10.3) is called the generalized Bellman equation for a process S in a discrete medium $\mathcal{M} = (X, \Gamma, L, A)$.

1.2 Analysis of Discrete Computational Media

The operation of modern computational systems is based on the idea of parallelism in implementing computational operations. There are two ways in which computations can be parallel [1.16-1.18]. The first involves parallel execution of computational operations on nonuniform calculating devices, that is, devices that differ in their functional characteristics. This method is widely used in modern supercomputers, such as CRAY-I, CRAY-II, CYBER-205, and *Elbrus*. The second method employs homogeneous CS. It is most effective for solving problems whose algorithms of solution allow for a representation in the form of many identical collections of operations (local subprograms), whose execution can be carried out in parallel. Such are the problems of linear algebra (the reversal and multiplication of matrices, or the solution of systems of linear equations), adaptive and recursive filtration, fast discrete Fourier transformations, the solution of systems of partial differential equations, the problems of sorting, of optimization on graphs, and of pattern recognition. The method is realized in a number of modern CS, such as ILLIAC-4, PS-2000, PS-3000, DAP, SLIP-4, and systolic array computers [1.17, 1.18].

The increase in the speed of operation of nonuniform CS is closely linked with perfecting the element basis, while the speed of homogeneous CS depends on the number N of calculating devices carrying out the executing local subprograms (homogeneous elementary computational systems).

In this section we will consider the mathematical models for analyzing the functioning of homogeneous CS. We will allow for the small parameter that arises quite naturally here, $h \in [0, 1]$, which is inversely proportional to the number N of homogeneous elementary computational systems. The collection of homogeneous

elementary CS and the system of data exchange channels is commonly known as a computational medium [1.17, 1.19]. Below we give a formal definition of a discrete computational medium and consider the formulations and methods of solution of problems dealing with the organization of calculations in such a medium. The conditions (formulated in Section 1.2.1) that a discrete computational medium must meet ensure the possibility of passage to the limit in the small parameter $h \to 0$ to the continuous model of a computational medium.

1.2.1 Discrete Computational Medium

Here we define the concept of a discrete computational medium, the architecture of such a medium, and the characteristic of the medium. The architecture and the characteristic define in a unique manner the linear operator A in the space of states of the medium.

We start by considering an ideal computational medium that completely fills the three-dimensional Euclidean space or a Euclidean plane (a flat structure). The effect that the finiteness of a medium has on the various processes taking place in a real computational medium with boundaries will be taken into account by stating the appropriate boundary conditions for each specific problem.

Let us introduce $\Omega = \Omega(\mathscr{L}, \mathscr{B})$, a translation-invariant lattice[2] in R^3. This lattice can be specified uniquely by fixing two sets of vectors, $\mathscr{L} = \{\alpha_1, \alpha_2, \alpha_3\}$ and $\mathscr{B} = \{\beta_1, \beta_2, \beta_3\}$, which form the base of the lattice. Note (see [1.20]) that the vectors α_1, α_2, and α_3 are always assumed to be linearly independent; the parallelepiped built on these vectors is known as the elementary, or primitive, cell of the lattice. The lattice's base vectors drawn from the origin, an apex of the primitive cell, are subjected to only one condition, namely, that

$$\beta_i = \sum_{j=1}^{3} r_i^j \alpha_j \Rightarrow 0 \leqslant r_i^j \leqslant 1, \quad j = 1, 2, 3, \quad i = 1, 2, \ldots, q. \tag{1.2.1.1}$$

The lattice $\Omega(\mathscr{L}, \mathscr{B})$ is formed by all the points of the space (the lattice points, or sites) that are the terminal points of the following set of vectors:

$$\Omega = \Big\{ x \,|\, x = \beta_i + \sum_{j=1}^{3} n_j \alpha_j, \quad n_j = 0, \pm 1, \pm 2, \ldots, \\ i = 1, 2, \ldots, q, \; j = 1, 2, 3 \Big\}. \tag{1.2.1.2}$$

[2] For the sake of definiteness we take the three-dimensional case. All further discussions can be applied to two-dimensional lattices in R^2 with no modifications.

From the definition of a lattice there directly follows the invariance of the lattice with respect to the discrete group of translations $T\ (\Omega)$ with generators T_{α_1}, $T_{,\alpha_2}$ and T_{α_3} : the lattice transforms into itself under an arbitrary translation of the form $T = k_1 T_{\alpha_1} + k_2 T_{\alpha_2} + k_3 T_{\alpha_3}$, where k_1, k_2, and k_3 are arbitrary integers, and T_{α_j} a translation along the vector α_j, $j = 1, 2, 3$.

Remark 1.2.1.1 A two-dimensional lattice with a pair of basis translation vectors α_1 and α_2 on plane R^2 can be defined in a similar manner. When base $\mathscr{B}$ consists of only one vector, the lattice is called the Bravais lattice [1.8]. This type of lattice plays and important role in solid state theory as the simplest model (two-dimensional or three-dimensional) of a crystal [1.21].

Let $M_\Omega = (X, \Gamma, L, A)$ be a discrete medium (see Section 1.1.10) for which the set of points X coincides with Ω, a lattice with a given pair $(\mathscr{L}, \mathscr{B})$, and $\Gamma = \Omega \times \Omega$. We introduce the following notation for the points of the medium: $\forall x \in X$, $x = (n, \beta)$, where $n = (n_1, n_2, n_3) \in \mathbb{Z}^3$, $\beta \in \mathscr{B}$, are the "coordinates" in expansion (1.2.1.2) for a vector terminating at x. The set of lattice sites $K\ (n) = \{x \mid x = (n, \beta),\ \beta \in \mathscr{B}\}$ is said to be the cell of the medium with number n.

The definitions that follow clarify the properties of the characteristic of connections L: $\Omega \times \Omega \to A$ of the points of medium M_Ω depending on whether we are considering "internal" connections, that is, connections between the points in the medium, $x = (n, \beta)$ and $x' = (n', \beta')$, belonging to the same cell ($n = n'$), or "external" connections, that is, connections between lattice sites belonging to different celles ($n \neq n'$).

Definition 1.2.1.1 The map f: $\mathbb{Z}^3 \to 2^{\mathbb{Z}^3}$ possessing the property $\forall n \in \mathbb{Z}^3$, $f\ (n) \ni 0$, is said to be the regulator of discrete medium M_Ω if the characteristic function of the menium, L: $\Gamma \to A$, satisfies the following condition:

$$\forall n \in \mathbb{Z}^3, \quad L\ ((n,\ \beta), \quad (n',\ \beta')) \neq \Im \Leftrightarrow n - n' \in f\ (n). \tag{1.2.1.3}$$

The set of vectors $f\ (n) = \{v_1, v_2, \ldots, v_i, \ldots, v_i \in \mathbb{Z}^3\}$ is said to be the regulator of the medium in cell $K\ (n)$; the cardinal number, or power, of this set, $p\ (n) = |\ f\ (n)\ |$, will be called the power of regulator f in the cell with number n.

To define a discrete computational medium let us consider the following properties of medium M_Ω:

(1) the homogeneity of the medium: there exists a map L_1: $\mathbb{Z}^3 \quad \mathscr{B} \times \mathscr{B} \to A$ such that $\forall\ ((n,\ \beta), \ (n',\ \beta')) \in \Gamma$,

$$L\ ((n,\ \beta), \ (n',\ \beta')) = L_1\ (n - n',\ \beta,\ \beta'); \tag{1.2.1.4}$$

(2) the local property of the connections: there is a positive constant ε such that $\forall n \in \mathbb{Z}^3$, $\forall v \in f(n)$,

$$\| v \| < \varepsilon, \tag{1.2.1.5}$$

with $\| v \|$ the norm of vector v, say, $\| v \| = \sum_{i=1}^{3} | v_i |$;

(3) the regularity of the connections: the map f is constant, $f(n) \overset{\text{def}}{=} V$ $\forall n \in \mathbb{Z}^3$, and the power of the regulator is finite: $| V | = p$, $p \in \mathbb{N}$.

Definition 1.2.1.2 A discrete computational medium is a medium $M_\Omega = \{\Omega, \Gamma, L, A\}$ for which the characteristic function (or the characteristic of connections) L possesses Properties (1)-(3), that is, the homogeneity of the medium, the local property of the connections, and the regularity of the connections.

Remark 1.2.1.2 The given definition of a computational medium satisfies the requirements usually considered in the design of homogeneous CS with a program-rearrangement structure [1.16, 1.17]. The effectiveness of paralleling computations in such systems depends on the homogeneity of the medium, the local property of the connections, and the regularity of the connections. By homogeneity of computational media we mean the similarity of the component structural elements (elementary processors, memory units, and data exchange circuits) in their functional characteristics, that is, the similarity of the elements proper (Property (1) with $n = n'$) and the similarity of the connections between the elements (Property (1) with $n \neq n'$). The local property of connections in a computational medium means the property of data exchange circuits to establish connections only between close (in a certain sense) elementary processors (Property (2)), while the regularity of connections in a CS means that the connections are recurrent and of a single type (Properties (1) and (3)).

Let us now fix the semi-ring A and the group T_α of translations in the lattice. Definition 1.2.1.2 implies that a discrete computational medium M_Ω is given if we have specified (a) the structural parameters of the medium: $\mathscr{B} \in \mathbb{R}^3$ (the base of lattice Ω), $q \in \mathbb{N}$ (the number of elements in the base), $V \subset \mathbb{Z}^3$ (the regulator of the medium), $p \in \mathbb{N}$ (the number of elements in the regulator), and ε (the extent to which the connections are local); (b) the characteristic of these parameters: the map L_1 defined in (1.2.1.3) and (1.2.1.4). Let us introduce the following notation: $(\mathscr{B}, q) = \mathscr{B}_q$ and $(V, p) = V_p$.

Definition 1.2.1.3 The pair $(\mathscr{B}_q, V_p) = \mathrm{Ar}$ is said to be the architecture of the discrete computational medium, and the map L_1: $V \times \mathscr{B} \times \mathscr{B} \to A$ is called the architecture characteristic.

The architecture of a medium and its characteristic define (see formula (1.1.10.3)) an endomorphism L_M: $A^X \to A^X$ in the space

of the states of the medium, $H_M = A^X$. In terms of coordinates (n, β) of the points of the medium, this endomorphism has the form

$$(L_M\varphi)(n, \beta) = \bigoplus_{v\in V}\bigoplus_{\beta'\in\mathscr{B}} L_1(v, \beta, \beta') \odot \varphi(n-v, \beta')$$

$$= \bigoplus_{v\in V}\bigoplus_{\beta\in\mathscr{B}} L_1(v, \beta, \beta') \odot \hat{T}_{-v}\varphi(n, \beta') \quad \forall\varphi\in A^x, \quad (1.2.1.6)$$

where $\hat{T}_{-v}$ is the endomorphism of the shift generated by the translation $T_{-v} \in T(\Omega)$ by the vector $-v$, with $(\hat{T}_{-v}(\varphi(n, \beta')) = \varphi(n - v, \beta'))$.

If we assume the semi-module $C_{\mathbb{O}}$ $(X = \mathbb{Z}^3 \times \mathscr{B})$ to be the semi-module $C_{\mathbb{C}}(\mathbb{Z}^3, A^q)$ of functions on $\mathbb{Z}^3$ with values in A^q, where A^q is the direct sum $A \oplus A \oplus \ldots \oplus A$, and define, for each $v \in V$, the endomorphism $\hat{L}(v)$ of semi-module A^q by the formula

$$(\hat{L}(v)g)(\beta) = \bigoplus_{\beta\in\mathscr{B}} L_1(v, \beta, \beta') \odot g(\beta') \quad \forall g\in A^q, \quad (1.2.1.7)$$

then from formula (1.2.1.6) for L_M as the endomorphism of semi-module $C_{\mathbb{C}}(\mathbb{Z}^3, A^q)$ we obtain the following representation $(L_M \overset{\text{def}}{=} \hat{L}_M)$:

$$(\hat{L}_M\varphi)(n) = \bigoplus_{v\in V} \hat{L}(v)\varphi(n-v). \quad (1.2.1.8)$$

Here $\varphi(n) \in A^q$, $n \in \mathbb{Z}^3$, the column vector $(\varphi_1(n), \varphi_2(n), \ldots, \varphi_q(n))^T$. In what follows the endomorphism $\hat{L}_M = (\hat{L}(v_0), \ldots, \hat{L}(v_{p-1}))$ will retain its meaning as the characteristic of the connections in medium M_Ω, with $\hat{L}(v_i)$ being A-valued q-by-q "matrices" $\hat{L}(v_i)$: $A^q \to A^q$ defined in (1.2.1.7); for the discrete computational medium M_Ω we will use the following notation: $M_\Omega = (T_\alpha, \mathrm{Ar}, \hat{L}_M, A)$. The generalized Bellman (evolution) equation with a right-hand side $\mathscr{F}_t$ for the process $s = \{s_t, t = 0, 1, \ldots, s_t \in C_{\mathbb{C}}(\mathbb{Z}^3, A^q)\}$ taking place in a discrete medium M_Ω has the form (cf. (1.1.10.2)-(1.1.10.3))

$$s_{t+1}(n) = \bigoplus_{v\in V} \hat{L}(v)s_t(n-v) \oplus \mathscr{F}_t(n), \quad (1.2.1.9)$$

$$s|_{t=0}(n) = s_0(n). \quad (1.2.1.10)$$

Here $n \in \mathbb{Z}^3$ is the number of the elementary cell of the medium, $s_t(n)$ is the A-valued vector describing the state of cell $K(n)$ at time t, the component $s_t^j(n)$, $j = 1, \ldots, q$, is the state of the jth point of the medium in cell $K(n)$, endomorphism $\hat{L}(v)$ is defined in

(1.2.1.7), s_0: $\mathbb{Z}^3 \to A^q$ is the initial state in process s, and $\mathscr{F}_t$: $\mathbb{Z}^3 \to A^q$ is a given family of functions.

We will study the steady-state Bellman equation corresponding to the endomorphism $\hat{L}_M$ given by (1.2.1.8), or

$$s = \hat{L}_M s \oplus \mathscr{F}, \tag{1.2.1.11}$$

where $\mathscr{F} \in C_{\mathbb{O}}(\mathbb{Z}^3, A^q)$ is a given function $\mathbb{Z}^3 \to A^q$ that is time independent. Solution s: $\mathbb{Z}^3 \to A^q$ to Eq. (1.2.1.11) is said to be a steady-state process in medium M_Ω.

Remark 1.2.1.3 For $q = 1$ the base $\mathscr{B} = \{\beta\}$ is called a one-point base. Without loss of generality we can always assume in this case that $\mathscr{B}$ consists of the zero vector $\beta = 0$. For $\beta \neq 0$ this can always be done by shifting the system of coordinates by vector β. For a one-point base the points of a medium coincide with the vertices of primitive cells. Correspondingly, Eqs. (1.2.1.9) and (1.2.1.11) are "scalar" equations with respect to the functions s: $\mathbb{Z}^3 \to A$. The scalar endomorphism $\hat{L}(v)$ given by (1.2.1.7) will be denoted in this case by $L(v)$.

The majority of application problems considered in this section and in Sections 1.3 and 1.4 lead to scalar Bellman equations ($q = 1$, $\mathscr{B} = \{0\}$). The general case ($q > 1$) requires employing the operator calculus of functions of "noncommutative" linear "operators" that assume values in appropriate semi-rings and will not be considered here (except in specific problems discussed in Sections 1.2.4.1 and 1.2.4.2).

1.2.2 Solution of the Bellman Equation in a Discrete Computational Medium

In this section we discuss the formulation of problems and explicit formulas for the resolving operators for the evolution and steady-state Bellman equations in a discrete computational medium. The calculations are based on Duhamel's theorem (Section 1), the translation invariance of the equations, and, partially, on the results of Section 5.

1.2.2.1 The Cauchy Problem

Suppose that $s_t \in C_{\mathbb{O}}(\mathbb{Z}^3, A)$ satisfies the inhomogeneous equation

$$s_{t+1}(x) = \bigoplus_{v \in V} L(v) \odot s_t(x - v) \oplus \mathscr{F}_t(x), \quad x \in \Omega = \mathbb{Z}^3,$$
$$t = 0, 1, \ldots, \tag{1.2.2.1}$$

with the initial condition

$$s_{t=0}(x) = s_0(x), \quad \mathbb{O} \leqslant S_0(x) \leqslant c_0, \quad c_0 = \text{const}, \tag{1.2.2.2}$$

where $\mathscr{F}_t$: $\mathbb{Z}^3 \to A^q$ is a given family of functions.

Theorem 1.2.2.1 *Let $\mathscr{F}_t$, $t = 0, 1, \ldots$, be a family of functions uniformly bounded in t, $\mathbb{O} \leqslant \mathscr{F}_t(x) \leqslant c$, $x \in \Omega$, $c =$ const, and let $L(v) \leqslant \mathbb{I}$ $\forall\, v \in V$. Then (1) problem (1.2.2.1), (1.2.2.2) has a unique bounded solution, and (2) solution $s_t(x)$ can be represented in the form*

$$s_t(x) = s_t^{(0)}(x) \oplus s_t^F(x), \quad x \in \Omega, \tag{1.2.2.3}$$

where $s_t^{(0)}(x)$ and $s_t^F(x)$ are the solution to the homogeneous equation (Eq. (1.2.2.1) with $\mathscr{F}_t(x) = \mathbb{O}$) with the initial condition (1.2.2.2) and, respectively, the solution to Eq. (1.2.2.1) with a zero initial condition; the two solutions are given by the following formulas:

$$s_t^{(0)}(x) = \bigoplus_{n \in \Sigma_t} \mathscr{L}_p(n) \odot s_0(x - \Lambda n), \tag{1.2.2.4}$$

$$s_t^F(x) = \bigoplus_{0 \leqslant \alpha \leqslant t} \bigoplus_{n \in \Sigma_\alpha} \mathscr{L}_{p,n} \odot \mathscr{F}_{\alpha-1}(x - \Lambda n), \tag{1.2.2.5}$$

where $n = (n_1, \ldots, n_p) \in \mathbb{Z}_+^p$, $\mathscr{F}_{-1} = \mathbb{O}$

$$\Lambda = \Lambda_{s \times p} = [v_1, v_2, \ldots, v_p], \quad \Sigma_t = \left\{ n \in \mathbb{Z}_+^p, \ \sum_{i=1}^p n_i = t \right\},$$

$$\mathscr{L}_p: \mathbb{Z}_+^p \to A, \quad \mathscr{L}_p(n) \overset{\text{def}}{=} L^{n_1}(v_1) \odot \ldots \odot L^{n_p}(v_p). \tag{1.2.2.6}$$

The proof of the theorem follows from Duhamel's theorem in explicit form for the resolving operator $\hat{R}_t$: $s_0 \to s_t$ of the problem

$$s_{t+1} = \hat{L}_M s_t, \quad s|_{t=0} = s_0, \quad t = 0, 1, \ldots, \tag{1.2.2.6a}$$

where $\hat{L}_M = \bigoplus_{v \in V} L(v)\, \hat{T}_{-v}$, and $\hat{T}_{-v}$ is the endomorphism of the "shift" $C_{\mathbb{D}}(\mathbb{Z}^3) \to C_{\mathbb{D}}(\mathbb{Z}^3)$ generated by the translation by vector $(-v)$: $\hat{T}_{-v}\varphi(x) = \varphi(x - v)$. Obviously,

$$\begin{aligned} \hat{R}_t \quad & \Big(\bigoplus_{v \in V} L(v) \odot \hat{T}_{-v} \Big)^t \\ & \bigoplus_{i_1} \bigoplus_{i_2} \ldots \bigoplus_{i_t} L(v_{i_1}) \odot L(v_{i_2}) \odot \ldots \odot L(v_{i_t}) \\ & \odot \hat{T}_{-v_{i_1}} * \hat{T}_{-v_{i_2}} * \ldots * \hat{T}_{-v_{i_t}}, \end{aligned}$$

where $\hat{T}_{-v} * \hat{T}_{-u}$ is a composition of endomorphisms, $\hat{T}_{-u} * \hat{T}_{-v} = \hat{T}_{-v} * \hat{T}_{-u} \quad \hat{T}_{-v-u}$, and hence $\hat{T}_{-u}^{n_1} * \hat{T}_{-v}^{n_2} = \hat{T}_{-n_1 u - n_2 v}$, $n_i \in \mathbb{Z}_+$,

$u, v \in \mathbb{Z}^3$. Therefore, if we allow for the idempotency of $\oplus$ and the commutativity of $\odot$, we can represent operator $\hat{R}_t$ in the form

$$\hat{R}_t = \bigoplus_{\substack{\sum_{i=1}^{p} n_i = t \\ n_i \in \mathbb{Z}_+}} \mathscr{L}_p(n) \odot \hat{T}_{-\sum_{i=1}^{p} n_i v_i}, \tag{1.2.2.7}$$

where

$$\mathscr{L}_p(n) = L^{n_1}(v_1) \odot \ldots \odot L^{n_p}(v_p),$$
$$L^{n_k}(v_k) = \underbrace{L(v_k) \odot L(v_k) \odot \ldots \odot L(v_k)}_{n_k \text{ times}}$$

($\mathscr{L}_p(n)$ is the A-valued analog of the action integral along the trajectory $\mu_{[0,t]}(n)$ that passes through the points $(\zeta, t = 0)$ and (x, t)), and

$$\mu_{[0,t]}(n) \overset{\text{def}}{=} \left\{ n, \ x - \zeta = \sum_{i=1}^{p} n_i v_i, \ \sum_{i=1}^{p} n_i = t \right\}$$

(see [1.2,1.3]).

Applying operator $\hat{R}_t$ (1.2.2.7) to s_0, we get (1.2.2.4). Formula (1.2.2.5) follows from Duhamel's theorem, in view of which the solution $s_t^F(x)$ to the problem

$$s_{t+1} = \hat{L}_M s_t \oplus \mathscr{F}_t, \ s_t(x) = \mathbb{0} \tag{1.2.2.8}$$

is given by the integral

$$s_t = \int\limits_{[0,t]}^{\oplus} W(t, \tau)\, d\tau \overset{\text{def}}{=} \bigoplus_{0 \leqslant \tau \leqslant t} W(t, \tau), \tag{1.2.2.9}$$

where $W(t, \tau) = \hat{R}(t - \tau)\, \mathscr{F}_{\tau-1}$. Employing (1.2.2.7) and introducing the variable $\alpha = t - \tau$, $0 \leqslant \alpha \leqslant t$, we arrive at (1.2.2.5). The uniqueness of the solution to problem (1.2.2.6) follows from formula (1.2.2.7) and the property $\oplus^2 = \oplus$, while the uniqueness of representation (1.2.2.5) for s_t^F was proved in Section 1.1.8.

Example. Let A be the semi-ring of the form $(R^1 \cup \{\pm\infty\}, \oplus = \max, \odot = +)$. Let $\mathscr{F}(x)$ be independent of t, or $\mathscr{F}_t(x) = \mathscr{F}(x)$. Finally, let $s_0(x) = \mathbb{0}$. Then (1.2.2.5) implies

$$s_t^F(x) = \max_{0 \leqslant \alpha \leqslant t} \max_{\substack{\sum_{i=1}^{p} n_i = \alpha \\ n_i \in \mathbb{Z}_+}} \left\{ \langle L, n \rangle + \mathscr{F}\left(x - \sum_{i=1}^{p} n_i v_i \right) \right\},$$

where $\langle L, n \rangle = \sum_{i=1}^{p} L(v_i)\, n_i$.

1.2.2.2 The Stabilization Cauchy Problem

This name is given to the following problem (see Section 1.0): find

$$\lim_{t \to +\infty} s(x, t) = s^*(s_0), \qquad (1.2.2.10)$$

where $s(x, t)$ is the solution to the Cauchy problem

$$s_{t+1} = \hat{L}_M s_t \oplus \mathcal{F}, \quad \mathbb{O} \leqslant \mathcal{F}(x) \leqslant c, \quad x \in \Omega \qquad (1.2.2.11)$$
$$s|_{t=0} = s_0.$$

Let us denote the resolving operator of problem (1.2.2.10), (1.2.2.11) by $\hat{B}_{s_0}$ (with s_0 fixed). Provided that $\lim\limits_{t\to\infty} \widetilde{L}_M^t = H^\infty$ and $\lim\limits_{t\to\infty} \hat{L}_M^{(t)} = H^*$ exist, where $\hat{L}_M^{(t)} \overset{\mathrm{def}}{=} \bigoplus\limits_{k=0}^{t} \hat{L}_M^k$, and that the limits are understood in the sense of strong convergence of the endomorphisms on the subspace of bounded functions taken from $C_{\mathbb{O}}(\Omega)$, we can define the operator $\hat{B}_{u_0}$: $C_{\mathbb{O}}(\Omega) \to C_{\mathbb{O}}(\Omega)$ by formula (1.5.1.8) (see Section 1.5):

$$\hat{B}_{s_0} \mathcal{F} = H^\infty s_0 \oplus H^* . \mathcal{F}. \qquad (1.2.2.12)$$

1.2.2.3 The Steady-state Bellman Equation

Let s satisfy the equation

$$s = \hat{L}_M s \oplus \mathcal{F}, \quad \mathbb{O} \leqslant \mathcal{F}(x) \leqslant c, \quad x \in \Omega = \mathbb{Z}^3. \qquad (1.2.2.13)$$

By virtue of the results arrived at in Section 1.1.8, the general solution to Eq. (1.2.2.13) is the sum of an arbitrary solution $\varphi_0(x)$ of the appropriate homogeneous equation and a special kind of solution s_F^* of the inhomogeneous equation that became known as the Duhamel solution. If for operator $\hat{L}_M$ the conditions (1.2.1.4) and (1.2.1.5) are met, s_F^* is a solution to the stabilization Cauchy problem at $s_0 = \mathbb{O}$, that is, $s_F^* = \hat{B}_{\mathbb{O}} . \mathcal{F} = H^* . \mathcal{F}$ (see Theorem 1.5.1.1).

Theorem 1.2.2.2 *The Duhamel solution to the generalized steady-state Bellman equation* (1.2.2.13) *has the form*

$$s_F^*(x) = \bigoplus_{n \in \mathbb{Z}_+^p} \mathcal{L}_p(n) \odot \mathcal{F}(x - \Lambda n), \qquad (1.2.2.14)$$

where $\mathcal{L}_p(n) = L^{n_1}(v_1) \odot L^{n_2}(v_2) \odot \ldots \odot L^{n_p}(v_p)$, *and* $\Lambda n = \sum\limits_{i=1}^{p} v_i n_i$.

Proof. From the definition of s_F^* and formula (1.2.2.5) it follows that

$$s_F^* = \lim_{t\to\infty} \hat{L}_M^{(t)} = \lim_{t\to\infty} \bigoplus_{0\leqslant\alpha\leqslant t} \bigoplus_{\sum_{i=1}^{p} n_i=\alpha} \mathscr{L}_p(n) \odot \mathscr{F}(x-\Lambda n)$$

$$= \bigoplus_{0\leqslant\alpha\leqslant\infty} \bigoplus_{\sum_{i=1}^{p} n_i=\alpha} \mathscr{L}_p(n) \odot \mathscr{F}(x-\Lambda n)$$

$$= \bigoplus_{\substack{n_i\in\mathbb{Z}_+\\ i=1,\ldots,p}} \mathscr{L}_p(n) \odot \mathscr{F}(x-\Lambda n).$$

Example. Let $A = ((R^1 \cup \{\pm\infty\})^n,\ \oplus = \min,\ \odot = +)$ and let $G(x, x_0)$ be the solution to Eq. (1.2.2.13) with the right-hand side $\mathscr{F} = \delta^{x_0}(x)$, where $\delta^{x_0}(x)$ is the "delta function" with respect to the scalar product $\langle\ ,\ \rangle_\oplus$ in $C_{\mathbb{O}}(\Omega)$,

$$\delta^{x_0}(x) = \begin{cases} \mathbb{1} & \text{if } x = x_0, \\ \mathbb{O} & \text{if } x \neq x_0 \end{cases}$$

($G(x, x_0)$ is the analog of the Green function). Then, by virtue of the linearity of problem (1.2.2.13), formula (1.2.2.14) yields the solution s_F^* for an arbitrary right-hand side, $s_F^*(x) = \min_{x_0} \{G(x, x_0) + \mathscr{F}(x_0)\}$. Say, for $x \in \mathbb{Z}$, if $V = \{1, -1\}$ $\forall v$, $L(v) = \mathbb{1} = 0$, then $G(x, x_0) = |x - x_0|$; but if $V = \{1, -1\}$, $L(1) = +1$, and $L(-1) = -1$, then $G(x, x_0) = x - x_0$.

1.2.2.4 The Steady-State Bellman Equation for a Restricted Discrete Computational Medium

A discrete computational medium M is said to be restricted if it contains an arbitrary but finite set of elementary cells, that is, $|\Omega| < \infty$. We introduce the notation $\widetilde{V} = \bigcup_{n\in\Omega} V(n)$, where (the reader will recall) $V(n)$ is the regulator of the medium at point n. A point $n \in \Omega$ for which $V(n) = \widetilde{V}$ is said to be an interior point of the medium, and $\mathring{\Omega} \overset{\text{def}}{=} \{n,\ V(n) = \widetilde{V}\}$ stands for the set of all interior points in the medium. The regulator of the interior points of the medium will be denoted by $\dot{V}$. Let us define the boundary points of medium M as the complement $\partial\Omega = \Omega \setminus \dot{\Omega}$; respectively, the regulator of the medium at a point $n \in \partial\Omega$ will be denoted by $V^\partial(n)$. Obviously, $V^\partial(n) \in \widetilde{V}$ $\forall n \in \partial\Omega$. For all $n \in \Omega$ we introduce the set $\Lambda(n) = \{v \in \widetilde{V},\ n - v \in \Omega\}$.

Definition 1.2.2.1 A restricted discrete computational medium is said to be regular if

$$V(n) = \Lambda(n) \quad \forall n \in \Omega. \tag{1.2.2.15}$$

Note that for $n \in \mathring{\Omega}$ condition (1.2.2.15) is certain to be met.

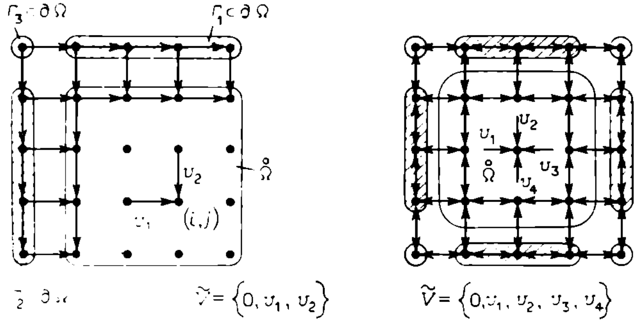


Fig. 1.2

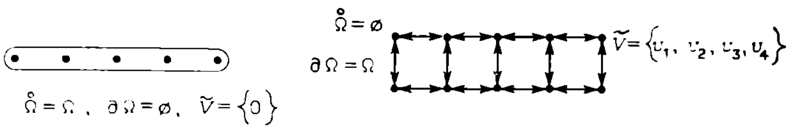


Fig. 1.3

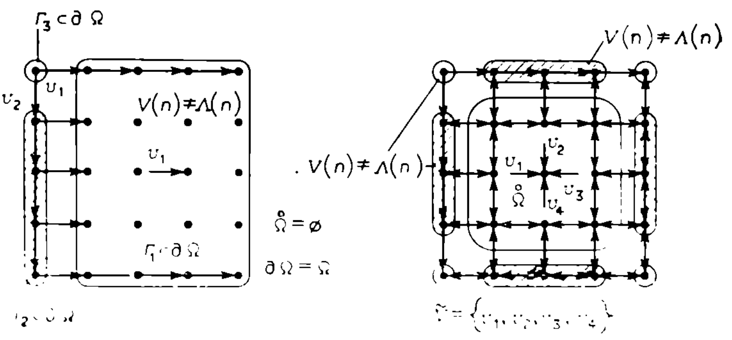


Fig. 1.4

By this definition, a regular restricted computational medium may have no boundary points and consist entirely of interior points. Examples of regular media are shown in Figures 2 and 3, while Figure 4 represents a nonregular (or irregular) medium.

Let us construct solutions to the steady-state Bellman equation in a regular discrete medium $M_\Omega = (\Omega, \mathrm{Ar}, \hat{L}_M, A)$, with $\mathrm{Ar} = (q = 1,\ B = \{0\};\ p,\ V = \mathring{V} \cup V^\partial)$:

$$s = \hat{L}_M s \oplus \mathcal{F} \iff \begin{cases} s(n) = \hat{L}_M(\mathring{V})\, s \oplus \mathring{\mathcal{F}}(n), & n \in \mathring{\Omega}, \\ \hat{L}_M(\mathring{V}) = \bigoplus\limits_{v \in \mathring{V}} L(v) \odot \hat{T}_{-v}; & \\ s(n) = \hat{L}_M(V^\partial)\, s \oplus \mathcal{F}^\partial(n), & n \in \partial\Omega, \\ \hat{L}_M(V^\partial) = \bigoplus\limits_{v \in V^\partial} L(v) \odot \hat{T}_{-v}. & \end{cases} \qquad (1.2.2.16)$$

For a regular discrete computational medium M satisfying the conditions that (i) Ω is a convex set and (ii) $L(v) \leqslant \mathbb{1}\ \ \forall\ v \in V$ we have the following

Theorem 1.2.2.3 *Let $\tilde{\mathcal{F}}$: $C_{\mathbb{C}}(\mathbb{Z}^3) \to A$ be a function obtained through a continuation of $\tilde{\mathcal{F}}$ with zero $\mathbb{O}$ outside Ω, and let $s^*_{\tilde{\mathcal{F}}}$ be the Duhamel solution to the generalized Bellman equation for a discrete (unrestricted) computational medium $M(\mathring{V})$ with regulator $\mathring{V}$,*

$$s^*_{\tilde{\mathcal{F}}}(n) = \hat{L}_M(\mathring{V})\, s^*_{\tilde{\mathcal{F}}}(n) \oplus \tilde{\mathcal{F}}(n), \quad n \in \mathbb{Z}^3. \qquad (1.2.2.17)$$

Then the restriction $s^*_{\tilde{\mathcal{F}}}|_\Omega$ of function $s^*_{\tilde{\mathcal{F}}}$ to Ω constitutes a solution to Eq. (1.2.2.16).

Proof. We will carry out the proof by using the discrete "continual" representation of solution $s^*_{\tilde{\mathcal{F}}}$ in the form

$$s^*_{\tilde{\mathcal{F}}}(x) = \bigoplus_{\zeta \in \mathbb{Z}^3} \ \bigoplus_{\tilde{\mu} \in Q(\zeta, x)} f(\tilde{\mu}) \odot \tilde{\mathcal{F}}(\zeta), \quad x \in \mathbb{Z}^3,$$

where

$$Q(\zeta, x) = \bigcup_{0 \leqslant \alpha \leqslant \infty} Q^\alpha(\zeta, x), \quad Q^\alpha(\zeta, x) = \bigcup_{n \in I(\alpha)} Q^\alpha_n(\zeta, x),$$

$$I(\alpha) = \Big\{ n = (n_i),\ i = 1, \ldots, p,\ \ x - \zeta = \sum_{i=1}^{p} n_i v_i,\ \ \sum_{i=1}^{p} n_i = \alpha \Big\},$$

with

$$Q^\alpha_n(\zeta, x) = \Big\{ \tilde{\mu}_\alpha(\zeta, x),\ \tilde{\mu}_\alpha(\zeta, x) = \Big(\zeta,\ \zeta + v_{i_1},\ \zeta + v_{i_1} + v_{i_2},\ \ldots,\ \zeta + \sum_{k=1}^{\alpha} v_{i_k}\Big),\ (i_1, i_2, \ldots, i_\alpha) \in J_\alpha \Big\}.$$

Here J_α is the set of permutations of length α, $\tilde{\mu}_\alpha\ (\zeta, x)$ is the trajectory of length α connecting points ζ and x, and $f(\tilde{\mu}_\alpha\ (\zeta,\ x)) = L^{n_1}(v_1) \odot \ldots \odot L^{n_p}(v_p)$ is the "contribution" of trajectory $\tilde{\mu}_\alpha (\zeta, x)$ to $s^*_{\tilde{\mathcal{F}}}$ in medium $M(\overset{\circ}{V})$. By virtue of condition (ii), the representation of the set of trajectories $Q(\zeta, x)$ that lead from ζ to x, and the definition of $\tilde{\mathcal{F}}$, there exists an $\alpha_0 < \infty$ such that

$$s^*_{\tilde{\mathcal{F}}}(x) = \bigoplus_{\zeta\in\Omega}\ \bigoplus_{0\leqslant\alpha\leqslant\alpha_0}\ \bigoplus_{n\in I(\alpha)}\ \bigoplus_{\tilde{\mu}\in Q^\alpha_n(\zeta,\, x)} f(\tilde{\mu}) \odot \mathcal{F}(\zeta),\ x \in \mathbb{Z}^3. \qquad (1.2.2.18)$$

Now suppose that $x \in \Omega$. Since $\forall\ \tilde{\mu} \in Q^\alpha_n\ (\zeta,\ x)$, we conclude that $f(\tilde{\mu})$ is independent of the interchange of the regulator vectors included in this route, and we have

$$s^*_{\tilde{\mathcal{F}}}(x) = \bigoplus_{\zeta\in\Omega}\ \bigoplus_{0\leqslant\alpha\leqslant\alpha_0} [(\bigoplus_{n\in I_\Omega(\alpha)} f(\mu) \odot \mathcal{F}(\zeta))$$

$$\oplus (\bigoplus_{n\in \bar{I}_\Omega(\alpha)} f(\mu) \odot \mathcal{F}(\zeta))], \qquad (1.2.2.19)$$

where $I_\Omega(\alpha)$ is the set of all $n \in I(\alpha)$ for which there is a trajectory $\mu_\Omega\ (\zeta,\ x) \in Q^\alpha_n\ (\zeta,\ x)$ all points of which lie in Ω, and $\bar{I}_\Omega(\alpha) = I(\alpha)\setminus I_\Omega(\alpha)$. By virtue of the condition that $\forall\, \bar{n} \in I(\alpha)$, there is an $n \in I_\Omega\ (\alpha)$ such that $\bar{n}_i \geqslant n_i$, $i = 1, \ldots, p$, whereby, by virtue of the axioms of metric and structure on the semi-ring A and the condition that $b \leqslant \mathbb{1} \Rightarrow \mathbb{1} \oplus b = \mathbb{1} \Rightarrow \forall a,\ \ a \oplus (a \odot b) = a$, we have

$$f(\bar{\mu}) = f(\mu) \odot L_1^{\bar{n}_1 - n_1} \odot \ldots \odot L_p^{\bar{n}_p - n_p},$$

where $\bar{\mu} \in Q^\alpha_{\bar{n}}\ (\zeta, x)$ and $\mu \in Q^\alpha_n\ (\zeta,\ x)$. Combining this with condition (ii), we obtain the formula

$$s^*_{\tilde{\mathcal{F}}}(x) = \bigoplus_{\zeta\in\Omega}\ \bigoplus_{0\leqslant\alpha\leqslant\alpha_0}\ \bigoplus_{n\in I_\Omega(\alpha)} f(\mu) \odot \mathcal{F}(\zeta),\ x \in \Omega,$$

which, by virtue of the condition of regularity imposed on M_Ω, can be transformed into

$$s^*_{\tilde{\mathcal{F}}}(x) = \bigoplus_{\zeta\in\Omega}\ \bigoplus_{\mu\in Q_\Omega(\zeta,\, x)} f(\mu) \odot \mathcal{F}(\zeta),\ x \in \Omega, \qquad (1.2.2.20)$$

where $Q_\Omega(\zeta,\ x)$ is the set of all trajectories in medium M_Ω that connect points ζ and x. The proof of the theorem is complete, since (1.2.2.20) and the continual representation (1.2.2.16) of solution s (see [1.22-24]) yield $s^*_{\tilde{\mathcal{F}}}\,|_\Omega = s$.

Example. Let us select a regular discrete medium M assuming that

(1) $\Omega = \{x \in \mathbb{Z}_+,\ 0 \leqslant x \leqslant N\},\ N \in \mathbb{Z}_+$;

(2) $\overset{\circ}{V}(x) = \{0,\ 1\},\ 1 \leqslant x \leqslant N,\ \ V^0(0) = \{0\}$;

(3) $A = (R^1 \cup \{\pm\infty\}, \oplus = \max, \odot = +)$;

(4) $L(1) = t_0$, $t_0 = \text{const} > 0$, $L(0) = \mathbb{O}$.

Then (1.2.2.16) assumes the form $s(x) = \max(s(x-1) + t_0, \mathcal{F}(x))$, $1 \leqslant x \leqslant N$, $s(0) = \mathcal{F}(0)$, where $\mathcal{F}(x)$, $0 \leqslant x \leqslant N$, is a given function, with $-\infty \leqslant \mathcal{F}(x) \leqslant c$. Computing the solution $s^*_{\tilde{\mathcal{F}}}(x)$, $x \in \mathbb{Z}$, via formula (1.2.2.14), and taking its restriction to Ω, we find that

$$s(x) = s^*_{\tilde{\mathcal{F}}}|_\Omega = \max_{0 \leqslant \zeta \leqslant x} \{\mathcal{F}(x - \zeta) + t_0\zeta\}.$$

Remark 1.2.2.1 In what follows, if the contrary is not stipulated, it will be assumed that the restricted discrete medium M_Ω, $|\Omega| < \infty$, is regular and satisfies conditions (i) and (ii) and, hence, the assertion of Theorem 1.2.2.3.

Remark 1.2.2.2 In the scalar case, if the contrary is not stipulated $L(v_0 = 0) = \mathbb{O}$.

1.2.3 Activity of a Homogeneous Multiprocessor Computational System

In this section we will describe the wavefront of a calculation process in a homogeneous computational system with an array architecture.

For a nonrestricted computational medium $M_\Omega \overset{\circ}{=} M^{\text{com}} = (T_\alpha, \text{Ar}, \hat{L}_M, A)$, we define the semi-ring A as the set $\{0, 1\}$ with commutative semi-group operations $\oplus = \min$ and $\odot = \max$ and with neutral elements $\mathbb{O} = 1$ and $\mathbb{1} = 0$, respectively. Let the base $\mathcal{B}$ of lattice Ω be of the one-point type, $\mathcal{B} = \{0\}$, and let the characteristic of connections, L_1, admit the value $\mathbb{1}$. The evolution Bellman equation for medium M^{com} has the form

$$s_{t+1}(n) = \bigoplus_{v \in V} s_t(n - v) = \min_{1 \leqslant i \leqslant p} \{s_t(n - v_i)\}. \qquad (1.2.3.1)$$

Suppose that the initial state of process s is

$$s_{t=0}(n) = s_0(n) = \begin{cases} \mathbb{O} \text{ if } n \notin \omega_0, \\ \mathbb{1} \text{ if } n \in \omega_0, \end{cases} \qquad (1.2.3.2)$$

where ω_0 is an arbitrary finite (compact) set in $\mathbb{Z}^3$. Knowing the solution to problem (1.2.3.1), (1.2.3.2), namely, knowing s_t, $t \in [0, T]$, $T > 0$ (see formulas (1.2.2.4) and (1.2.2.5)), we can find the set of the points of the medium, or

$$\omega_{[0,\, T]}(V_p) \overset{\text{def}}{=} \bigcup_{0 \leqslant t \leqslant T} \Phi_t, \qquad (1.2.3.3)$$

where Φ_t is the wavefront of process s at time $t \in [0, T]$, $\Phi_t \overset{\text{def}}{=} \{n, s_t(n) = \mathbb{1}\}$. We call $\omega_{[0,T]}(V_p)$ the activity range of the computational medium M^{com} over the time interval $[0, T]$.

This model of a discrete medium M^{com} and a process $\{s_t, t = 0, 1, \ldots\}$ that satisfies (1.2.3.1) and (1.2.3.2) enable analyzing, on the basis of formula (1.2.3.3), the functioning activity of a homogeneous multiprocessor computational system controlled by a flow of data as a function of the architecture of the system. Within the framework of the model suggested here, a point $n \in \Omega$ of the medium corresponds to an elementary processor in a homogeneous computational system, and the vector $v \in V$ defines the channel of data exchange in the direction from processor $n - v$ to processor v. The state $s_t(n)$ at time $t = 0, 1, \ldots$ admits only two values: $s_t(n) = \mathbb{1}$ means that processor n is executing calculations, or is active, and $s_t(n) = \mathbb{0}$ means that processor n is in the "wait" state. The operation of a multiprocessor CS (see [1.4-1.7]) presupposes that each elementary processor functions according to its own individual (local) program, in which the existence is stipulated of operators of data exchange between neighboring processors along connections v allowed for in the system ($v \in V$). It is also assumed that processor n will become active at time $t + 1$ when it receives the results of calculations carried out by the neighboring processors, which have terminated their operation at time t. The latter condition is satisfied thanks to Eq. (1.2.3.1). For a given connection architecture $\mathrm{Ar} = (V_p, \{0\})$, the set $\omega_{[0,T]}(V_p)$ stipulates the processors in the computational system that have terminated their operation, that is, were active at times $t \in [0, T]$. At time zero, the set of active processors, ω_0 (1.2.3.2), is determined by the flow of data into the system at $t = 0$.

A model of an array processor. The previous model makes it possible to describe the operation of a homogeneous computational system with the so-called array structure (an array processor) [1.4, 1.5, 1.7].

A discrete computational medium $M^{\mathrm{com}} = (T_\alpha, \mathrm{Ar}, \hat{L}_M, A)$ corresponding to an array processor [1.5, 1.6] is determined by the following values of its parameters: Ω is a two-dimensional lattice consisting of N times N (N a positive integer) primitive cells; the unit vectors T_α in R^2, with $\alpha_1 = (0, 1)$ and $\alpha_2 = (1, 0)$, are the generators of the translation group (see Figure 2); the architecture of the array processor is fixed by the one-point base $\mathcal{B} = \{0\}$ and the regulator $V_p(n)$, which assumes the following values:

(a) $p = 2$ and $V = \{\alpha_1, \alpha_2\}$ for $n \in \mathring{\Omega}$;

(b) for $n \in \partial\Omega = \bigcup_{i=1}^{3} \Gamma_i$, where

$$\Gamma_1 = \{(i, 1), i = 2, \ldots, N\},$$
$$\Gamma_2 = \{(1, j), j = 2, \ldots, N\}, \ \Gamma_3 = \{(1, 1)\}, \qquad (1.2.3.3')$$

we have $p = 1$ and $V^{\partial}(n) = \{\alpha_2\}$ if $n \in \Gamma_1$, $p = 1$ and $V^{\partial}(n) = \{\alpha_1\}$ if $n \in \Gamma_2$, and $p = 0$ and $V^{\partial}(n) = \varnothing$ if $n \in \Gamma_3$; $\hat{L}_M$ and A are the same as in Section 1.2.1.

The calculation process s_t, $t = 0, 1, \ldots$, in such a medium satisfies the following system of equations (see (1.2.3.1) and (1.2.3.2)):

$$\begin{aligned} s_{t+1}(n) &= \min\{s_t(n-\alpha_1),\ s_t(n-\alpha_2)\} \\ &= \min\{s_t(i-1,\ j),\ s_t(i,\ j-1)\}, \\ & n \in \mathring{\Omega},\ n = (i,\ j),\quad i,\ j \in 2, \ldots, N, \end{aligned} \tag{1.2.3.4}$$

$$s_{t+1}(n) = \min\{\min_{v \in V^\partial}\{s_t(n-v)\},\ \mathcal{F}_t(n)\},\ n \in \partial\Omega, \tag{1.2.3.5}$$

$$s_t|_{t=0}(n) = s_0(n), \tag{1.2.3.6}$$

where $\mathcal{F}_t$ is a function $\partial\Omega \to A$ given for every value of t.

The "source" $\mathcal{F}_t(n)$ in the right-hand side of (1.2.3.5) describes the interaction of the processors lying at the boundary of the medium with the external memory controlling the calculation process. The value $\mathcal{F}_t(n) = \mathbb{1}$ means that at time t the processor $n \in \partial\Omega$ is allowed either to transform the results of its calculations to the memory or to receive new data from the external memory in accordance with the calculation algorithm.

For example, for the case where $\mathcal{F}_t(n) = \mathbb{1}$ and $s_t|_{t=0}(n) = \delta(n - n_0)$, where

$$\delta(n-n_0) = \begin{cases} \mathbb{1} & \text{if } n = n_0 = (1,\ 1), \\ \mathbb{0} & \text{if } n \neq n_0, \end{cases} \tag{1.2.3.7}$$

we can find the activity range of the array processor, $\omega_{[0,T]}(V_p)$, by (a) solving Eq. (1.2.3.5), employing Duhamel's theorem, (c) solving Eq. (1.2.3.4), and (d) employing (1.2.3.3). From the explicit formulas (1.2.2.4) and (1.2.2.5) for the solutions, we readily get

$$\omega_{[0,\ T]}(V_p) = \begin{cases} \{(i,j),\ i+j = T,\ i \geqslant 1\} & \text{if } 0 \leqslant T \leqslant 2N-2, \\ \varnothing & \text{if } T > 2N-2. \end{cases}$$

Remark 1.2.3.1 Let the function $\mathcal{F}_t(n)$, $n \in \partial\Omega$, be such that $\forall t\ \mathcal{F}_t(i,\ 1) = \mathcal{F}_t(i,\ N)$ and $\mathcal{F}_t(1,\ j) = \mathcal{F}_t(N,\ j)$, $i = 1, \ldots, N$, $j = 1, \ldots, N$. In this case, replacing Eq. (1.2.3.5) with the periodic boundary conditions $\forall t \geqslant 0\ \ s_t(n + N\alpha_1) = s_t(n)$ and $s_t(n + N\alpha_2) = s_t(n)$, we obtain the following Bellman equation on the torus T^2:

$$s_{t+1}(n) = \min\{\min_{v \in V}\{s_t(n-v)\},\ \mathcal{F}_t(n)\}.$$

In the common linear case, a similar problem for a crystal with Born-von Kármán boundary conditions has been discussed in [1.12].

1.2.4 The Effectiveness of Parallel Programs

The concept of a generalized Bellman equation applied to discrete computational media makes it possible, at least in principle, to solve the problem of designing homogeneous computational systems from the stand-

point of the highest effectiveness of realization in such computational media of the executed programs for a given class of problems. On the basis of the solution to the Bellman equation, we estimate the effectiveness of parallel programs for matrix multiplication, *LU*-expansion of a matrix into two triangular matrices, and solution of systems of linear equations.

The problem of designing a computational system lies in the choice of the architecture corresponding to the discrete computational medium M_Ω that provides an optimal value for the quality criterion $\gamma(\mathrm{Ar})$ of the architecture $\mathrm{Ar} = (V_p, \mathscr{B}_q)$. It has been established that all such criteria are linear in spaces with values in certain semirings (cf. Section 1.3). For example,

$$\gamma(\mathrm{Ar}) = \sum_{x \in K} \nu(x) \min_{P \in \mathscr{P}(\mathrm{Ar},\, x)} T(P) \quad \text{or}$$

$$\gamma(\mathrm{Ar}) = \max_{x \in K} \min_{P \in \mathscr{P}(\mathrm{Ar},\, x)} T(P),$$

where K is the given class of problems being solved, $T(P)$ is the time of execution of program $P \in \mathscr{P}(\mathrm{Ar}, x)$, $\mathscr{P}(\mathrm{Ar}, x)$ is the set of all programs that solve the problem $x \in K$ with a selected architecture of the computational system, and $\nu(x)$ is the relative frequency of solution of problem x by the user. Thus, the problem of designing a CS has been reduced to estimating the time $T(P)$ of execution of the problem $P \in \mathscr{P}(\mathrm{Ar}, x)$ that solves problem $x \in K$ with the given architecture $\mathrm{Ar}(M_\Omega)$ of the discrete computational medium M_Ω.

Suppose that we have selected a homogeneous computational system for solving problems of a given class K, that is, suppose that the discrete computational medium $M_\Omega = (T_\alpha, \mathrm{Ar}, \hat{L}_M, A) \overset{\mathrm{def}}{=} M^{\mathrm{com}}$ is given.

A program $P \in \mathscr{P}(\mathrm{Ar}, x)$ is understood to be a parallel program constituting a collection of local programs[3], which, generally speaking, depend on parameter $n \in \Omega$, or $P = \{P^{\mathrm{loc}}(n),\ n \in \Omega\}$, and interact via the operators of interprocessor data exchange operators "PUT" and "GET". Let us describe these operators. To this end, we use the given architecture $\mathrm{Ar} = (V_p, \mathscr{B}_q)$ to define the set

$$V^s = V \cup V^- \cup \{0\}, \tag{1.2.4.1}$$

where V is the regulator in medium M^{com}, and $V^- = \{-v, v \in V\}$, and we fix the numbering of the elements in V^s, or $V^s = \{v_0, v_1, \ldots, v_l, \ldots, v_s\}$, where $v_0 = 0$, $l = 0, 1, \ldots, s$, $s = 2p'$, $p' < p$.

The numbering of the elements in set V^s fixes the possible operators of interprocessor data exchange, precisely, the operator GET l, $\langle I \rangle$; and the operator PUT l, $\langle I \rangle$; where $\langle I \rangle$ is the identifier of the

[3] Here $P^{\mathrm{loc}}(n)$ is understood to be a program for a single-processor computer.

variable, and $l = 1, \ldots, s$. Local programs will be described via the Dijkstra controlling structures IF . . . FI and DO . . . OD. The parametrization of a local program $P^{loc}(n)$, $n \in \Omega$ is carried by the operands $\mathfrak{G}i$, with $\mathfrak{G}i \overset{\text{def}}{=} n_i$, where n_i is the ith coordinate in $n = (n_1, n_2, n_3) \in \Omega$.

Let us refine the statement of the problem of estimating the time $T(P)$ for execution of the program P. Let P be a given program taken from the set of all the programs $\mathcal{P}(\text{Ar}, x)$ that solve problem x with the selected architecture of the computational system, $P = \{P^{loc}(n), n \in \Omega\}$, and let $t(P^{loc}(n))$ be the time of termination of the local program on processor n, while t_0 is the time at which the processors in the computational system begin to be loaded with the local programs. Then

$$T(P) = \max_{n \in \Omega} \{t(P^{loc}(n))\} - t_0. \tag{1.2.4.2}$$

It appears that the function $n \to t(P^{loc}(n))$, which, by virtue of (1.2.4.2), determines the time $T(P)$ of execution of the entire program P, may be effectively calculated on the basis of the solution to the generalized steady-state Bellman equation (1.2.1.11) in the space of functions with values in the semi-ring $A = (R^1_+, \oplus = \max, \odot = +)$.

The discrete computational medium $M_\Omega \overset{\text{def}}{=} M^P = (T^P_\alpha, \text{Ar}^P, \hat{L}^P_M, A)$ that fixes the coefficients in this equation is determined by program P in the following manner:

(1) $T^P_\alpha = T_\alpha$;

(2) $V^P(n) \subseteq V^s$, where V^s was defined in (1.2.4.1), $n \in \Omega$; and

(3) the base $\mathcal{B}^P$ and the characteristic of connections $\hat{L}_M = \hat{L}^P_1(v_0), \hat{L}^P_1(v_2), \ldots, \hat{L}^P_1(v_s)$ are determined uniquely by the text of local program $P^{loc}(n)$ and the time of execution of elementary operations in the program.

Let us now describe the procedure of constructing the base $\mathcal{B}^P$ and the endomorphism $\hat{L}^P_M$. By $\text{Tr}\, P^{loc}(n) = \{\zeta_i, 1 \leqslant i \leqslant M\}$ we denote the trace of $P^{loc}(n)$, a quantity that represents the sequence of elementary operations carried out by processor n operating according to program $P^{loc}(n)$. Let the function f: $\text{Tr}\, P^{loc}(n) \to R^1_+$ specify the time of execution of elementary operations in program $P^{loc}(n)$. Next we define the sets $\forall v_l \in V^P \setminus \{0\}$

$$\begin{aligned} &K^-(n, v_l) = \{i, (\zeta_i = \text{GET}\, l, \langle I \rangle;) \vee (\zeta_i = \text{PUT}\, l, \langle I \rangle;) \\ &l = 1, \ldots, s\}, \\ &K^-(n) = \bigcup_{v_l \in V^P \setminus \{0\}} K^-(n, v_l), \; K^+(n) = \{i, i-1 \in K^-(n)\}, \\ &K(n) = \{1\} \cup K^-(n) \cup K^+(n) \cup \{Q+1\}, \end{aligned} \tag{1.2.4.3}$$

with $Q = |\operatorname{Tr} P^{\mathrm{loc}}(n)|$, $n \in \Omega$. On set $K(n)$ we fix the ordering relation for $\hat{\imath}_0$, where $\hat{\imath}_0\colon K(n) \to \{1, 2, \ldots, |K(n)|\}$, precisely, $\hat{\imath}_0(i) < \hat{\imath}_0(i') \Leftrightarrow i < i' \quad \forall i' \in K(n)$, $\forall i \in K(n)$, $i \neq i'$. The pair $(K, \hat{\imath}_0)$ specifies the base $\mathscr{B}^P$ via formula (1.2.1.1):

$$\beta(m) = \sum_{j=1}^{3} r^j(m)\,\alpha_j, \quad r^j(m) = (1 - i)/i,$$
$$i = 1, \ldots, q, \quad q = |K|, \tag{1.2.4.4}$$

where $m \in K$, with $i = i_0(m)$.

Remark 1.2.4.1 Since data is exchanged between processors in a pair if and only if one of the two is executing operator GET and the other, operator PUT, we have

$$|K^-(n, v_l)| = |K^-(n + v_l, -v_l)| \quad \forall n \in \mathring{\Omega}, \ l = 1, \ldots, s,$$
$$\forall i \quad \zeta_{k_i} \neq \zeta_{k_i'}, \ k_i \in K^-(n, v_l), \ k_i' \in K^-(n + v_l, -v_l),$$

that is, if $\zeta_{k_i} = \mathrm{PUT}$, then $\zeta_{k_i'} = \mathrm{GET}$ and, respectively, if $\zeta_{k_i} = \mathrm{GET}$, then $\zeta_{k_i'} = \mathrm{PUT}$. The characteristic of connections, $\hat{L}_M$, in medium M^P, or $\hat{L}_M = (\hat{L}(0), \hat{L}(v_l), l = 1, \ldots, s)$, will be defined by the following formulas:

$$v = v_0, \quad L_{ij}(0, n) = \begin{cases} \displaystyle\sum_{i'=k_i}^{k_{i+1}-1} f(\zeta_{i'}) & \text{if } i = j + 1, \\ \mathbb{O} & \text{if } i \neq j + 1, \end{cases} \tag{1.2.4.5}$$

$$v = v_l, \quad L_{ij}(v_l, n) = \begin{cases} t_n \text{ if } i = i_r + 1, \ j = i_r', \\ r = 1, \ldots, |K^-(n, v_l)| \\ i_r \in \hat{I}(n, v_l), \ i_r' \in \hat{I}(n + v_l, -v_l), \ l = 1, \ldots, s, \\ \mathbb{O} \qquad \text{otherwise,} \end{cases}$$

where the set

$$\hat{I}(n, v_l) \overset{\mathrm{def}}{=} \{\hat{\imath}_0(x), \ x \in K^-(n, v_l)\} \tag{1.2.4.6}$$

is assumed to be ordered in such a way that $i_{r_1} < i_{r_2} \Leftrightarrow r_1 < r_2$. In (1.2.4.5), $t_n = f(\zeta)$, with $\zeta = \mathrm{GET}$ or $\zeta = \mathrm{PUT}$, is the time of execution of interprocessor data exchange operations.

Remark 1.2.4.2 It follows from the definition of a parallel program P and formulas (1.2.4.3) and (1.2.4.6), that the medium M^P corresponding to this program is generally not homogeneous ($\mathscr{B}_q = \mathscr{B}_q(n)$, $L = L(n, V(n))$, $n \in \Omega$). Yet it is possible (see below) to partition the lattice Ω into subregions $\{\Omega_\alpha, \alpha = 1, 2, \ldots, \alpha_0\}$ in such a way that the corresponding media are homogeneous and,

hence, to reduce the solution of problem (1.2.4.2) to the consecutive solution of the Bellman equations in the homogeneous media $\mathcal{M}^P(\Omega_\alpha)$, $\alpha \in 1, \ldots, \alpha_0$.

Let us use specific examples to illustrate this approach to estimating the time of program execution. Suppose that we have fixed the architecture of a homogeneous CS. Let us assume, for the sake of definiteness, that we have chosen the architecture of the process examined in Section 1.2.3 (see p. 000), that is, $\mathrm{Ar} = (V_P, \mathscr{B} = \{0\})$, where $V_P = \{\alpha_1, \alpha_2\} = \{(1, 0), (0, 1)\}$. Then

$$V^- = \{(-1, 0), (0, -1)\},$$
$$V^s = \{v_0 = (0, 0), v_1 = (-1, 0), v_2 = (0, -1),$$
$$v_3 = (1, 0), v_4 = (0, 1)\}.$$

Hence, the admissible operators of interprocessor data exchange in the language of program description are GET l, $\langle I \rangle$; and PUT l, $\langle I \rangle$; with $l = 1, 2, 3, 4$. Let us now consider in greater detail the simple example of multiplication of two square (N-by-N) matrices A and B.

1.2.4.1 A Parallel Program of Matrix Multiplication

Let us decompose matrix A in its columns and matrix B in its rows:

$$A = [a_1, a_2, \ldots, a_N], \; a_i = (a_i^1, \ldots, a_i^N)^{\mathrm{T}}, \quad i = 1, \ldots, N;$$
$$B = [b_1, b_2, \ldots, b_N]^{\mathrm{T}}, \; b_i = (b_i^1, \ldots, b_i^N), \quad i = 1, \ldots, N.$$

We represent the product of these two matrices, $C = A \times B$, in the form

$$C = A \times B = \sum_{i=1}^{N} a_i \times b_i. \qquad (1.2.4.7)$$

Let us assume that the initial data in the program, the matrices A and B, are stored in the external memory modules: matrix A in the left memory modules in rows and matrix B in the right memory modules in columns (see Figure 1.5). Matrix $C = A \times B$ obtained as a result of executing the program $P_{A \times B}$ is stored in the memory of the elementary processors in the CS (the matrix element C_{ij} is stored in the memory of processor $n = (i, j)$, $i, j = 1, \ldots, N$).

Remark 1.2.4.3 In the given case, to simplify the program we assume that (a) the "dimensionality of the array processor" (the number of primitive cells in the computational medium M^{com}) and the dimensionality of the problem being solved (the size of matrices A and B) coincide, and (b) on a part of the boundary of the medium, $\partial\Omega = \{(i, j) \mid i = N \text{ or } j = N\}$ (see Figure 1.5), the operations of data transfer PUT 3, $\langle I \rangle$ and PUT 4, $\langle I \rangle$, respectively, are suppres-

sed. In general, a special operator is required to establish the boundary of an array consisting of processors on which data transfer operations will be suppressed.

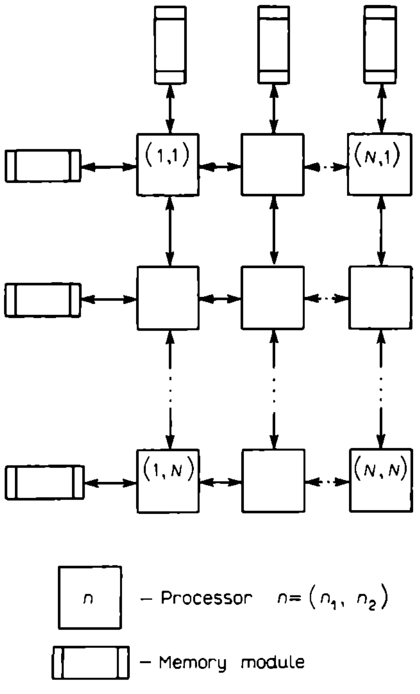


Fig. 1.5

Here is the text of a possible local program for solving problem (1.2.4.7).

Program $P^{\mathrm{loc}}_{A\times B}\ (n)$:

$$
\begin{array}{ll}
C := \varnothing; & I := 1; \\
\mathrm{DO}\ \mathrm{I} \Leftarrow \mathrm{N} \Rightarrow & \mathrm{GET}\ 1,\ \mathrm{A}; \\
 & \mathrm{GET}\ 2,\ \mathrm{B}; \\
 & \mathrm{PUT}\ 3,\ \mathrm{A}; \\
 & \mathrm{PUT}\ 4,\ \mathrm{B}; \\
 & C := C + A * \mathrm{B}; \\
 & I := I + 1 \\
\mathrm{OD}. &
\end{array}
\qquad (1.2.4.8)
$$

Using formulas (1.2.4.3)-(1.2.4.5), we can find the base and characteristic of connection of the medium $M_{\Omega}^{P_{A\times B}} = (T_{\alpha}^{P}, \mathrm{Ar}^{P}, L_{M}^{P}, A)$, where $P = \{P^{\mathrm{loc}}(n),\ n \in \Omega\}$.

Remark 1.2.4.4 In this case, as the text of the program (1.2.4.8) shows, the local program is independent of the number of the processor $n \in \Omega$. Let us now define

$$\begin{aligned}
&\mathrm{Tr}\, P^{\mathrm{loc}}\ (n) = \{\zeta_1 = \text{“=”};\ \ \zeta_2 = \text{“=”};\ \ \zeta_3 = \text{“}\Leftarrow\text{”};\\
&(\zeta_{10I-6} = \text{“GET 1”};\ \ \zeta_{10I-5} = \text{“GET 2”};\\
&\zeta_{10I-4} = \text{“PUT 3”};\ \ \zeta_{10I-3} = \text{“PUT 4”};\\
&\zeta_{10I-2} = \text{“*”};\ \ \zeta_{10I-1} = \text{“+”};\ \ \zeta_{10I} = \text{“=”};\\
&\zeta_{10I+1} = \text{“+”};\ \ \zeta_{10I+2} = \text{“=”};\\
&\zeta_{10I+3} = \text{“}\Leftarrow\text{”};)\ I = 1, \ldots, N\},\ \ Q = 10N + 3.
\end{aligned}$$

Let us define the function f: $\mathrm{Tr}\, P^{\mathrm{loc}}\ (n) \to R_{+}^{1}$, or

$$\begin{aligned}
&f(\text{“=”}) = t_{=},\ f(\text{“}\Leftarrow\text{”}) = t_{\leqslant},\ f(\text{“+”}) = t_{+},\ f(\text{“*”}) = t_{*},\\
&f(\text{“GET 1”}) = f(\text{“GET 2”}) = f(\text{“PUT 3”}) = f(\text{“PUT 4”})\\
&\qquad = t_{n}, \qquad (1.2.4.9')
\end{aligned}$$

and the set (see (1.2.4.3))

$$\begin{aligned}
&K^{-}(n,\ v_1) = \{10I\text{-}6,\ I = 1, \ldots, N\},\\
&K^{-}(n,\ v_2) = \{10I\text{-}5,\ I = 1, \ldots, N\}, \qquad (1.2.4.9'')\\
&K^{-}(n,\ v_3) = \{10I\text{-}4,\ I = 1, \ldots, N\}\ \ (v_3 = -v_1)\\
&K^{-}(n,\ v_4) = \{10I\text{-}3,\ I = 1, \ldots, N\}\ \ (v_4 = -v_2)
\end{aligned}$$

and, respectively, the sets (see (1.2.4.6) and (1.2.4.3))

$$\begin{aligned}
&\hat{I}(n,\ v_1) = \{5I - 3,\ I = 1, \ldots, N\},\ \hat{I}(n,\ v_2) = \{5I - 2,\\
&\qquad I = 1, \ldots, N\},\\
&\hat{I}(n,\ v_3) = \{5I - 1,\ I = 1, \ldots, N\},\ \hat{I}(n,\ v_4) = \{5I,\ I = 1, \ldots,\\
&\qquad N\}, \qquad (1.2.4.9''')\\
&K^{-}(n) = \{10I - j,\ j = 3, 4, 5, 6,\ I = 1, \ldots, N\},\\
&K^{+}(n) = \{10I - j,\ j = 2, 3, 4, 5,\ I = 1, \ldots, N\},\\
&K(n) = \{1\} \cup \{10I - j,\ j = 2, 3, 4, 5, 6,\ I = 1, \ldots, N\}\\
&\qquad \cup \{10N + 4\}.
\end{aligned}$$

Thus, the number of elements in base $\mathscr{B}^{P}$ is $|\,K(n)\,| = 5N + 2$. Using formulas (1.2.4.3)-(1.2.4.5), we write the characteristic of

connections, $\hat{L}_M$, element by element ($i,\ j = 1, \ldots, 5N + 2$):

$$L_{ij}(v_0) = \begin{cases} 2t_= + t_\leqslant, & i = j+1,\ j = 1, \\ t_n, & i = j+1,\ j = 5I - l, \\ & l = 0,\ 1,\ 2,\ 3,\ I = 1, \ldots, N, \\ 2t_= + 2t_+ + t_* + t_\leqslant, & i = j+1,\ j = 5I+1, \\ & I = 1, \ldots, N, \\ \mathbb{O} & \text{otherwise} \end{cases}$$

$$L_{ij}(v_1) = \begin{cases} t_n, & i = 5I - 2,\ j = 5I - 1,\ I = 1, \ldots, N, \\ \mathbb{O} & \text{otherwise}, \end{cases}$$

$$L_{ij}(v_2) = \begin{cases} t_n, & i = 5I - 1,\ j = 5I,\ I = 1, \ldots, N, \\ \mathbb{O} & \text{otherwise} \end{cases}$$

$$L_{ij}(v_3) = \begin{cases} t_n, & i = 5I,\ j = 5I - 3,\ I, \ldots, N, \\ \mathbb{O} & \text{otherwise}, \end{cases} \tag{1.2.4.10}$$

$$L_{ij}(v_4) = \begin{cases} t_n, & i = 5I + 1,\ j = 5I - 2,\ I = 1, \ldots, N, \\ \mathbb{O} & \text{otherwise}. \end{cases}$$

The text of program $P^{\text{loc}}(n)$ is used to determine the state vector $s_n = (s_1, \ldots, s_{5N+2})$ of medium $M^{P_{A\times B}}$: $s_1(n)$ is time of initiation of $P^{\text{loc}}(n)$, $s_2(n)$ is the time when the first iteration starts, $s_{5I-2}(n)$ is the end of operation GET 1, A, $s_{5I-1}(n)$ is the time of termination of operation GET 2, B, $s_{5I}(n)$ is the time of termination of operation PUT 3, A, $s_{5I+1}(n)$ is the time of termination of operation PUT 4, B, and $s_{5I+2}(n)$ is the time of termination of the Ith operation of a cycle of the program, $I = 1, \ldots, N$.

Assertion 1.2.4.1 *The time of execution $T(P)$ of program $P_{A\times B}$ (1.2.4.8) for solving problem (1.2.4.7) of multiplication of matrices A and B is given by the formula*

$$T(P_{A\times B}) = \max_{n\in\Omega} \{s_{5N+2}(n)\} - t_0, \tag{1.2.4.10$'$}$$

where $s_{5N+2}(n)$ is a component of solutions of the steady-state Bellman equation for medium $M^{P_{A\times B}}$ with a right-hand side $\mathcal{F}$ equal to

$$\mathcal{F}_i(n) = \begin{cases} t_1(n), & i = 1, \\ \mathbb{O} & i > 1, \end{cases} \quad i = 1, \ldots, 5N + 2, \tag{1.2.4.11}$$

where $t_1(n)$ is the time of termination of the loading of program $P^{\text{loc}}_{A\times B}(n)$ into processor $n \in \Omega$.

Proof. The proof follows from (1.2.4.2) and the algorithm for the construction of process s in the discrete computational medium $M^{P_{A\times B}}$, (1.2.4.8)-(1.2.4.10′).

Let us find the solution to problem (1.2.4.10′), (1.2.4.11). To this end we consider a discrete computational medium $\widetilde{M}_\Omega = (\widetilde{T}_\alpha, \widetilde{\text{Ar}}, \hat{L}_{\widetilde{M}}, A)$ for which the lattice Ω and the translation group $\widetilde{T}_\alpha$ are the same as for an array processor (see p. 000). We select the architecture and the characteristic of connections as follows: $\widetilde{\text{Ar}}\ ((p = 4,\ \widetilde{V}),\ (\mathscr{B}, q = 2))$, $\widetilde{V} = \{v_0 = (0, 0),\ v_1 = (1, 0),\ v_2 = (0, 1),\ v_3 = (1, 1)\}$, and $\hat{L}_{\widetilde{M}} = (\hat{L}(v_0), \hat{L}(v_1), \hat{L}(v_2), \hat{L}(v_3))$, $\hat{L}(v_i) = t_p^2 \odot \sigma_i$, $i = 0, 1, 2, 3$, where

$$\sigma_0=\begin{bmatrix}\mathbb{0} & \mathbb{0}\\ \mathbb{1} & \mathbb{0}\end{bmatrix},\quad \sigma_1=\begin{bmatrix}\mathbb{1} & \mathbb{0}\\ \mathbb{0} & \mathbb{0}\end{bmatrix},$$
$$\sigma_2=\begin{bmatrix}\mathbb{0} & \mathbb{0}\\ \mathbb{0} & \mathbb{1}\end{bmatrix},\quad \sigma_3=\begin{bmatrix}\mathbb{0} & \mathbb{1}\\ \mathbb{0} & \mathbb{0}\end{bmatrix}. \tag{1.2.4.12}$$

Let $\hat{R}_{\widetilde{M}}$ be the resolving operator of the problem

$$u=\hat{L}_{\widetilde{M}}u \oplus \mathscr{F},\quad \mathbb{0}\leqslant \mathscr{F}(x) < c,\quad u \in C_{\mathbb{0}}(\Omega, A^q). \tag{1.2.4.13}$$

Assertion 1.2.4.2 *The solution* $s \in C_{\mathbb{0}}(\Omega, A^q)$, $q = 5N + 2$, *to the Bellman equation for medium* $M^{P_{A\times B}}$ *with the right-hand side* $\mathscr{F}$ *taken from* (1.2.4.11) *is determined uniquely by the solution to problem* (1.2.4.13) *for medium* $\widetilde{M}_\Omega$; *in particular, for the component* $s_{5N+2}(n)$ *of vector* $s(n)$ *the following representation holds true*:

$$s_{5N+2}(n) = t_p^{2N-1} \odot t_2^N \odot [\hat{T}_{v_2}\ \hat{\mathbb{1}}]\,\hat{R}_{\widetilde{M}}(\Lambda \hat{R}_{\widetilde{M}})^{N-1} f^{(0)}(n), \tag{1.2.4.14}$$

where $\hat{R}_{\widetilde{M}}$ *is the resolving operator of problem* (1.2.4.13), $\Lambda = \begin{bmatrix}\hat{T}_{v_2} & \hat{\mathbb{1}}\\ \hat{T}_{v_3} & \hat{T}_{v_1}\end{bmatrix}$, *and* $f^{(0)}(n) = (F_1^{(0)}(n) \odot t_p, F_1^{(0)}(n + v_1) \odot t_p)^T$, *with* $F_1^{(0)}(n) = t_1(n) + 2t_= + t_\leqslant$, *where parameters* $t_=$, t_+, $t_\leqslant$, t_p, *and* $t_1(n)$ *have been defined earlier, and* $t_2 = 2t_= + 2t_+ + t_* + t_\leqslant$.

Proof. We will prove this assertion in two stages. First in view of the cyclicity of matrix $L(v)$, $v \in V$ (see Eq. (1.2.4.10)), the initial Bellman equation (1.2.1.11) with right-hand side $\mathscr{F}$ (1.2.4.11) can be transformed to an equivalent one-parameter family of equations of

the type

$$W^{(I)}(n) = \hat{L}W^{(I)} \oplus F^{(I-1)}, \quad \text{with } I \text{ the parameter},$$
$$I = 1, \ldots, N, \; n \in \Omega^{\mathrm{com}}, \quad (1.2.4.15)$$

$$F^{(I)}(n) = (W_6^{(I)}, \mathbb{O}, \mathbb{O}, \mathbb{O}, \mathbb{O}, \mathbb{O})^{\mathrm{T}}, \quad I = 1, \ldots, N-1,$$
$$F^{(0)}(n) = (t_1(n) + 2t_= + t_\leqslant, \mathbb{O}, \mathbb{O}, \mathbb{O}, \mathbb{O}, \mathbb{O})^{\mathrm{T}}, \qquad (1.2.4.16)$$

where $\hat{L} = \{L(v_k),\ k = 0, 1, 2, 3, 4\}$, $L_{ij}(v_k) = L_{i+1, j+1}(v_k)$, $i, j = 1, 2, 3, 4, 5, 6$ and $L_{ij}(v_k)$ is defined in (1.2.4.10), while the components of solution $W^{(I)}$ are expressed in terms of solution $s \in A^{5N+2}$ thus: $W_k^{(I)}(n) = s_{5I+k-4}(n)$, $k = 1, 2, 3, 4, 5, 6$. Next, the system (1.2.4.15) can be reduced by the method of elimination to a system of two equations in the components $W_2^{(I)}(n)$ and $W_4^{(I)}(n)$:

$$W_2^{(I)}(n) = F_1^{(I-1)} \odot t_n \oplus W_2^{(I)}(n-v_1) \odot t_n^2 \oplus W_4^{(I)}(n-v_3) \odot t_n^2,$$
$$(1.2.4.17)$$
$$W_4^{(I)}(n) = F_1^{(I-1)}(n+v_1) \odot t_n \oplus W_2^{(I)}(n) \odot t_n^2 \oplus W_4^{(I)}(n-v_2) \odot t_n^2,$$

with

$$s_{5I+2} = W_4^{(I)} \oplus \hat{T}_{-v_2} W_2^{(I)} \odot c_1, \; c_1 = t_n \odot t_2. \qquad (1.2.4.17')$$

Introducing the matrices σ_i, $i = 0, 1, 2, 3$ (see (1.2.4.12)) and the notation

$$f^{(I)}(n) = (F_1^{(I)}(n) \odot t_n, \; F_1^{(I)}(n+v_1) \odot t_n), \; I = 0, \ldots, N-1,$$
$$(W_2^{(I)}(n), \; W_4^{(I)}(n))^{\mathrm{T}} = u^{(I)}(n) \in A^2,$$

we write system (1.2.4.17) in the form

$$u^{(I)}(n) = \hat{L}_{\tilde{M}} u^{(I)}(n) \oplus f^{(I-1)}(n), \; I = 1, \ldots, N. \quad (1.2.4.18)$$

Employing (1.2.4.16), we can easily express $f^{(I)}$ in terms of $f^{(I-1)}$ thus:

$$f^{(I)} = t_n \odot c_1 \odot \begin{bmatrix} \hat{T}_{-v_2} & \mathbb{1} \\ \hat{T}_{-v_3} & \hat{T}_{-v_1} \end{bmatrix} \hat{R}_{\tilde{M}} f^{(I-1)}, \qquad (1.2.4.19)$$

where $\hat{R}_{\tilde{M}}$ is the resolving operator of problem (1.2.4.13), and the action of the endomorphism matrix $\Lambda = [\hat{\lambda}_{ij}] = \begin{bmatrix} \hat{T}_{-v_2} & \mathbb{1} \\ T_{-v_3} & \hat{T}_{-v_1} \end{bmatrix}$ on the A-valued vector $u(n) = (u_1, u_2)^{\mathrm{T}}$ is defined by the relationship $(\Lambda u)_i(n) = \bigoplus_{j=1} \hat{\lambda}_{ji} u_j(n)$. Then (1.2.4.19) yields $f^{(N-1)} =$

$c_2^{N-1} \odot (\Lambda \hat{R}_{\widetilde{M}})^{N-1} f^{(0)}$, where $c_2 = t_n \odot c_1$. Hence, $u^{(N)} = \hat{R}_M f^{(N-1)} = c_2^{N-1} \odot \hat{R}_M (\Lambda \hat{R}_{\widetilde{M}})^{N-1} f^{(0)}$. By virtue of (1.2.4.17′) we finally establish that $s_{5N+2} = t_n^{2N-1} \odot t_2^N \odot [\hat{T}_{-v_2} \mathbb{1}] \hat{R}_{\widetilde{M}} (\Lambda \hat{R}_{\widetilde{M}})^{N-1} f^{(0)}$.

Assertion 1.2.4.3 *The resolving operator* $\hat{R}_{\widetilde{M}}$ *of problem* (1.2.4.13) *has the form*

$$\hat{R}_{\widetilde{M}} \begin{pmatrix} \mathcal{F}_1 \\ \mathcal{F}_2 \end{pmatrix} (n) = \max_{i \geqslant 0,\, j \geqslant 0} \begin{pmatrix} 2t_n (i+j+1) + \max \{ \mathcal{F}_1 (n-(i+1) v_1 - j v_2), \\ \mathcal{F}_2 (n-(i+1) v_1 - (j+1) v_2 \} \\ 2t_n (i+j+1) + \max \{ \mathcal{F}_1 (n - i v_1 - j v_2), \\ \mathcal{F}_2 (n - i v_1 - (j+1) v_2 \} \end{pmatrix}. \quad (1.2.4.19')$$

Proof. We have $\hat{R}_{\widetilde{M}} = \lim_{t \to \infty} \hat{R}_t$, where $\hat{R}_t$ is the resolving operator of the Cauchy problem $v_{t+1} = \hat{L}_{\widetilde{M}} v_t \oplus \mathcal{F}$, $v|_{t=0} = \mathbb{O}$. As in the scalar case, Duhamel's theorem yields

$$v_t = \int\limits_{0 \leqslant \tau \leqslant t}^{\oplus} \hat{R}_{t-\tau} \mathcal{F} = \bigoplus_{0 \leqslant \tau \leqslant t} \Big(\bigoplus_{i=1}^{p} \sigma_i \hat{T}_{-v_i} \Big)^{t-\tau} \mathcal{F}, \quad p = |V|.$$

Using the obvious "commutation" relations for the A-valued matrices σ_i, $i = 0, 1, 2, 3$, $\sigma_0^2 = \sigma_3^2 = \hat{\mathbb{O}}$, $\sigma_2^2 = \sigma_2$, $\sigma_1^2 = \sigma_2$, $\sigma_i \sigma_j = \hat{\mathbb{O}}$, $(i, j) \in \{(0, 2), (1, 0), (1, 2), (2, 1), (2, 3), (3, 1)\}$, $\sigma_0 \sigma_1 = \sigma_0$, $\sigma_3 \sigma_2 = \sigma_3$, $\sigma_0 \sigma_3 = \sigma_2$, $\sigma_1 \sigma_3 = \sigma_3$, where $\hat{\mathbb{O}}$ is the null matrix, the operator methods developed in [1.13], and, in particular, formulas taken from [1.14], we obtain

$$\hat{R}_t \begin{pmatrix} \mathcal{F}_1 \\ \mathcal{F}_2 \end{pmatrix} = \max_{0 \leqslant \alpha \leqslant t} 2t_n^{\alpha} \odot \max_{i+j=\alpha-1} \begin{pmatrix} \max \{ \mathcal{F}_1 (n-(i+1) v_1 - j v_2), \\ \mathcal{F}_2 (n-(i+1) v_1 - (j+1) v_2 \} \\ \max \{ \mathcal{F}_1 (n - i v_1 - j v_2), \\ \mathcal{F}_2 (n - i v_1 - (j+1) v_2) \} \end{pmatrix}.$$

Passing to the limit as $t \to \infty$, we get (1.2.4.19′).

Corollary 1.2.4.1 *The time* $T(P)$ *of execution of program* $P_{A \times B}$ *is equal to*

$$T(P) = t_1(1, 1) + N(8t_n + t_2) + t_1 - 2t_n, \quad t_1 = 2t_= + t_\leqslant, \\ t_2 = t_1 + 2t_+ + t_*,$$

where $t_1(1, 1)$ *is the time it takes to load* $P_{A \times B}^{\mathrm{loc}}(n)$ *by an optimal loader into processor* $n = (1, 1)$.

1.2.4.2 A Parallel Program of LU-expansion

Statement of the problem. Given a homogeneous multiprocessor computational system, specify a parallel program and estimate the time of its execution for solving the problem of representing an N-by-N matrix A in the form

$$A = L \times U$$

(the LU-expansion), where L is the lower triangular matrix with units on the principal diagonal, and U is the upper triangular matrix. Suppose that we have decomposed the matrix L in its columns and matrix U in its rows, as in Section 1.2.4.1:

$$A = \sum_{i=1}^{N} L_i U_i. \tag{1.2.4.20}$$

The calculation scheme of the LU-expansion is as follows:

$$A^{(1)} = A, \quad A^{(i+1)} = A^{(i)} - L_i U_i, \quad i = 1, \ldots, N-1,$$

where

$$\begin{aligned} U_i &= \{0, 0, \ldots, 0, a_{ii}^{(i)}, a_{ii+1}^{(i)}, \ldots, a_{iN}^{(i)}\}, \\ L_i &= \frac{1}{a_{ii}^{(i)}} \{0, 0, \ldots, 0, a_{ii}^{(i)}, a_{ii+1}^{(i)}, \ldots, a_{iN}^{(i)}\}^{\mathrm{T}}. \end{aligned} \tag{1.2.4.21}$$

Here is the text of a possible local program for solving problem (1.2.4.20). It is assumed that matrix A and the results of calculating matrices L and U are stored in the memory of elementary processors of the computational system.

Program $P_{LU}^{\mathrm{loc}}(n)$:

```
IF ϖ1 ⇐ ϖ2 ⇒ Q = ϖ1 − 1;
#  ϖ2 ⇐ ϖ1 ⇒ Q = ϖ2 − 1;
FI;
I = 1;
DO  I ⇐ Q ⇒ GET 2, U; PUT 4, U;
             GET 1, L; PUT 3, L;                        (1.2.4.22)
             A = A − L * U; I = I + 1;
OD;
IF ϖ1 > ϖ2 ⇒ PUT 4, A; U = A;
#  ϖ1 = ϖ2 ⇒ R = 1/A; PUT 4, R; U = A;
#  ϖ1 < ϖ2 ⇒ GET 2, R; PUT 4, R; L = A * R; PUT 3, L;
FI.
```

The discrete computational medium $M^{P_{LU}}$ for program (1.2.4.22) is determined by the general formulas (1.2.4.3)-(1.2.4.6). Omitting

the intermediate computations, which are similar to those in program $P_{A\times B}$, we note only the form of the set $K(n)$ (see (1.2.4.3)):

$$K(n) = \{1\} \cup \{10I - J,\ J = 0, 1, 2, 3, 4,\ I = 1, \ldots, Q\}$$
$$\cup \{10Q + J,\ J = 6, 7, 8, 10, 11,\ n_1 < n_2\}$$
$$\cup \{10Q + J,\ J = 6, 8, 9, 10,\ n_1 = n_2\}$$
$$\cup \{10Q + J,\ J = 6, 7, 8,\ n_1 > n_2\},$$

where $Q = \min(n_1, n_2) - 1$, with (n_1, n_2) the coordinates of an elementary processor in the computational system. Hence, the power of base $\mathscr{B}(n)$ of the local program $P_{LU}^{\mathrm{loc}}(n)$ is $|K(n)| = 5\min(n_1, n_2) + \operatorname{sgn}(n_2 - n_1)$ and, therefore, medium $M^{P_{LU}}$ is nonhomogeneous. Note that at $n_1 = n_2$ a new element $10Q + 6$ is included in the set $K(n)$. This makes it possible, as in Assertion 1.2.4.1, to reduce the Bellman equation corresponding to the nonhomogeneous medium $M^{P_{LU}}$ for a steady-state process s: $\Omega \to A^{|K(n)|}$ to a system of Bellman equations each of which is related to a nonhomogeneous medium. It proves possible (see Section 1.2.4.1) to break up the lattices of these media into subregions in such a manner that the discrete computational media corresponding to this separation are homogeneous and, hence, regular (see p. 000). Let us consider the restricted nonhomogeneous media M_R and M_G whose lattices allow for the separations $\Omega_R = \bigcup_{j=1}^{4} \Omega_j$ and $\Omega_G = \Omega_2 \cup \Omega_4$, respectively, as shown in Figure 1.6. Here

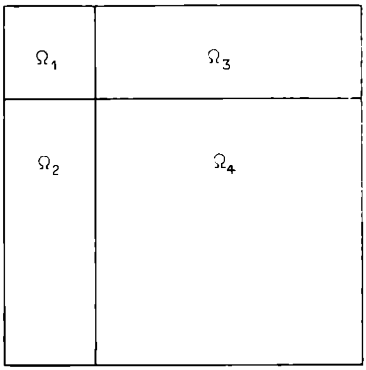


Fig. 1.6

$\Omega_1 = \{(I, I)\}$, $\Omega_2 = \{(I, j),\ I + 1 \leqslant j \leqslant N\}$,
$\Omega_3 = \{(i, I),\ I + 1 \leqslant i \leqslant N\}$, $\Omega_4 = \{(i, j),\ I + 1 \leqslant i, j \leqslant N\}$,

and I is a parameter (the number of iterations in $P_{LU}^{\mathrm{loc}}(n)$), with $I = 1, \ldots, Q$. The homogeneous computational media corresponding to these separations are $M_j(R) = (\Omega_j, \mathrm{Ar}_j(R), \hat{L}_j(R), A)$, $j = 1, 2, 3, 4$ and $M_j(G) = (\Omega_j, \mathrm{Ar}_j(G), \hat{L}_j(G), A)$, $j = 2, 4$. These can be found in the following way (the dependence of these media on parameter I is not given explicitly):

(1) the Ω_j are shown in Figure 1.6;

(2) the dimensionality of the base, q_j, and the regulators V_j are, respectively,

$$q_j(R) = 4,\ j = 1, 2;\quad q_j(R) = 3,\ j = 3, 4;\quad q_2(G) = 2,\ q_4(G) = 4;$$
$$V_j(R) = \{v_0\},\ j = 1, 3;\quad V_j(R) = \{v_0, v_2, v_4\},\ j = 2, 4;$$
$$V_2(G) = \{v_0\},\quad V_4(G) = \{v_0, v_3, v_4\},\quad V_j^{\partial}(R) = \{v_0, v_2\},\ j = 2, 4;$$
$$V_j^{\partial}(R) = \varnothing,\ j = 1, 3;\quad V_4^{\partial}(G) = \{v_0, v_1\},\quad V_2^{\partial}(G) = \varnothing;$$

(3) the characteristics of connection of the points of the medium, $\hat{L}(v)$, have the form

$$\hat{L}_1^R(v_0) = t_1 \odot [\varepsilon^{21}] \oplus t_n \odot [\varepsilon^{32}] \oplus t_4 \odot [\varepsilon^{43}],$$
$$\hat{L}_2^R(v_0) = t_n \odot [[\varepsilon^{21}] \oplus [\varepsilon^{32}]] \oplus t_5 \odot [\varepsilon^{43}],$$
$$\hat{L}_2^R(v_2) = t_n \odot [\varepsilon^{31}],\quad \hat{L}_2^R(v_4) = t_n \odot [\varepsilon^{22}],$$
$$\hat{L}_3^R(v_0) = t_n \odot [\varepsilon^{21}] \oplus t_4 \odot [\varepsilon^{32}],$$
$$\hat{L}_4^R(v_0) = t_n \odot [[\varepsilon^{21}] \oplus [\varepsilon^{32}]],$$
$$\hat{L}_4^R(v_2) = t_n \odot [\varepsilon^{31}],\quad \hat{L}_4^R(v_4) = t_n \odot [\varepsilon^{22}],$$
$$\hat{L}_2^G(v_0) = t_n \odot [\varepsilon^{21}],\quad \hat{L}_4^G(v_0) = t_n \odot [[\varepsilon^{21}] \oplus [\varepsilon^{32}]] \oplus t_2 \odot [\varepsilon^{43}];$$
$$\hat{L}_4^G(v_1) = t_n \odot [\varepsilon^{31}],\quad \hat{L}_4^G(v_3) = t_n \odot [\varepsilon^{22}].$$

Here we have introduced the notation $[\varepsilon^{i_0 j_0}]_{m\times n}$ for the m-by-n matrix with elements

$$(\varepsilon^{i_0 j_0})_{ij} = \begin{cases} \mathbb{1} & \text{if } i = i_0 \wedge j = j_0, \\ \mathbb{O} & \text{if } i \neq i_0 \vee j \neq j_0, \end{cases}\quad t_1 = 2t_{\leqslant} + 2t_{=} + t_{-},$$
$$t_2 = 2t_{=} + t_{*} + t_{-} + t_{+} + t_{\leqslant},$$
$$t_1 = t_{/} + t_{=},\ t_4 = t_{=},\ t_5 = t_{*} + t_{=},$$

where $t_{\leqslant}$, $t_{=}$, t_{-}, t_{+}, t_{*}, and $t_{/}$ are the times of execution of the elementary operations in program (1.2.4.22). Let us now write the systems of Bellman equations for the nonhomogeneous media M_R and M_G, respectively:

$$n \in \mathring{\Omega}_{\gamma},\quad u(n) = \hat{L}_{\gamma}(\mathring{V})\, u(n) \oplus \mathcal{F}_{\gamma}^{(I)}(n) \oplus \hat{L}_{\beta\to\gamma} u, \tag{1.2.4.23}$$

$$n \in \mathring{\Omega}_{\beta},\quad u(n) = \hat{L}_{\beta}(\mathring{V})\, u(n) \oplus \mathcal{F}_{\beta}^{(I)}(n). \tag{1.2.4.24}$$

$$u \in \partial\Omega_{\beta},\ u(n) = \hat{L}_{\beta}(V^{\partial})\, u(n) \oplus \mathcal{F}_{\beta}^{(I)}(n) \oplus \hat{L}_{\gamma\to\beta} u. \tag{1.2.4.25}$$

Here $(u, \gamma, \beta) \in \{(R^{(I)}, 1, 2); (R^{(I)}, 3, 4); (G^{(I)}, 2, 4)\}$, $I = 1, \ldots, Q$, where $R^{(I)}$ and $G^{(I)}$ are the states of media M_R and M_G, respective-

ly, and the operators $\hat{L}_{\beta\to\gamma}$ and $\hat{L}_{\gamma\to\beta}$ describe the "interaction" of homogeneous media and have the form

$$\begin{aligned}
&\hat{L}_{2\to 1} = t_n \odot [\varepsilon^{31}], \quad \hat{L}_{1\to 2} = t_n \odot [\varepsilon^{22}],\\
&\hat{L}_{4\to 3} = t_n \odot [\varepsilon^{21}], \quad \hat{L}_{3\to 4} = t_n \odot [\varepsilon^{21}],\\
&\hat{L}_{4\to 2} = t_n \odot [\varepsilon^{21}], \quad \hat{L}_{2\to 4} = t_n \odot [\varepsilon^{21}].
\end{aligned} \tag{1.2.4.26}$$

The functions $\mathscr{F}_j(R)$, $j = 1, 2, 3, 4$, and $\mathscr{F}_j(G)$, $j = 2, 4$, are given by the following formulas:

$$\mathscr{F}_j^{(I)}(R)(n) = \left(\begin{Bmatrix} t_1^{(I)}(n) + t_1 & \text{if } I = 1 \\ G_4^{(I-1)}(n) & \text{if } I > 1 \end{Bmatrix}, \underbrace{0, \ldots, 0}_{q_j(R)-1 \text{ time}}\right), \quad j = 1, 2, 3, 4,$$

$$\mathscr{F}_2^{(I)}(G)(n) = (R_4^{(I)}(n), 0), \quad \mathscr{F}_4^{(I)}(G)(n) = (R_3^{(I)}(n), 0, 0).$$

Assertion 1.2.4.4 *The solution* $s\colon \Omega \to A^{|K(n)|}$ *to the initial Bellman equation can be expressed in terms of solution* $R^{(I)}$ *and* $G^{(I)}$ *of system* (1.2.4.23)-(1.2.4.25) *according to the following formulas:*

$$\begin{aligned}
&s_{5I+i+1}(n) = R_i^{(I)}(n), \quad i = 1, 2, 3,\\
&s_{5Q+i+1}(n) = R_i^{(Q)}(n), \quad i = 1, 2, 3, 4, \; n_1 = I,\\
&n_2 \geqslant I; \; i = 1, 2, 3, \quad n_1 > I, \; n_2 = I,\\
&s_{5I+j+3}(n) = G_j^{(I)}(n), \; j = 1, 2, 3, 4, \; n_1 \geqslant I, \; n_2 > I,\\
&s_{5Q+i+4}(n) = G_i^{(Q)}(n), \; i = 1, 2, \; n_1 = I, \; n_2 > I.
\end{aligned} \tag{1.2.4.27}$$

The validity of (1.2.4.27) is verified by direct calculations that allow for (1.2.4.3)-(1.2.4.6).

Corollary *The time* $T(P_{LU})$ *of termination of operation of program* P_{LU} *is equal to* $T(P_{LU}) = \max\limits_{n\in\Omega} s_{|K(n)|}(n) - t_0$, where

$$s_{|K(n)|}(n) = \begin{cases} G_2^{(Q)}(n) & \text{if } n_2 > n_1, \\ R_4^{(Q)}(n) & \text{if } n_2 = n_1, \\ R_3^{(Q)}(n) & \text{if } n_2 < n_1. \end{cases} \tag{1.2.4.28}$$

In view of the last formula, to calculate $T(P_{LU})$ we need only find the solutions to system (1.2.4.23)-(1.2.4.26). By direct calculation it can easily be demonstrated that the operators of "interaction" of media M_R and M_G satisfy the following "commutation" relations:

$$\hat{L}_{\beta\to\gamma}\hat{L}_{\beta}(V^{\theta}) = \hat{\mathcal{O}}, \quad \hat{L}_{\beta\to\gamma}\hat{L}_{\gamma\to\beta} = \hat{\mathcal{O}} \quad \forall(\gamma, \beta). \tag{1.2.4.29}$$

In view of (1.2.4.29), for a fixed pair (γ, β) the solution $u_\gamma(n)$, $n \in \mathring{\Omega}_\gamma$, can be found from the equation

$$u(n) = \hat{L}_\gamma(\mathring{V})\, u(n) \oplus \mathscr{F}_\gamma^{(I)}(n) \oplus \hat{L}_{\beta\to\gamma}\mathscr{F}_\beta^{(I)}. \qquad (1.2.4.30)$$

Substitution of the solution of Eq. (1.2.4.30) into (1.2.4.25) leads to an equation for $u(n)$, $n \in \Omega_\beta$. This solution satisfies the Bellman equation in a restricted homogeneous medium:

$$u(n) = \hat{L}_\beta u(n) \oplus \Phi_\beta^{(I)}(n), \qquad n \in \Omega_\beta, \qquad (1.2.4.31)$$

where

$$\Phi_\beta^{(I)}(n) = \begin{cases} \mathscr{F}_\beta^{(I)}(n), & n \in \mathring{\Omega}_\beta, \\ \mathscr{F}_\beta^{(I)}(n) \oplus \hat{L}_{\gamma\to\beta}u_\gamma, & n \in \partial\Omega_\beta. \end{cases} \qquad (1.2.4.32)$$

In view of the properties of the characteristics of connections, which determine the operators $\hat{L}_\beta$ for $\beta \in \{2, 4\}$ in the media M_R and M_G (see p. 000), all the components $(u_\beta^{(I)}(n))_i$, $i = 1, 2, \ldots, q(\beta)$, of the solution $u_\beta^{(I)}$ to Eq. (1.2.4.31) are expressed in terms of the function $f_\beta^{(I)}(n)$, which for each $\beta \in \{2, 4\}$ satisfies the one-dimensional Bellman equation on the half-line:

$$f^{(I)}(x) = \max\,(t_n + f^{(I)}(x+1),\ g^{(I)}(x)), \quad x > I + 1;$$
$$f^{(I)}(I+1) = g^{(I)}(I+1). \qquad (1.2.4.33)$$

Here the variable $x = n_2$ if $u_\beta^{(I)} = R^{(I)}$ and $x = n_1$ if $u_\beta^{(I)} = Q^{(I)}$, where $(n_1, n_2) = n$ is the "number" of the processor in the computational system, while the function $g^{(I)}(x)$ is obviously determined by the right-hand side of (1.2.4.32). The solution to Eq. (1.2.4.33) was obtained in Section 1.2.2 (see Theorem 1.2.2.3). Substituting this solution into formulas (1.2.4.27)-(1.2.4.28) and assuming that $\max\,(t_i, t_n) = t_i$, $i = 3, 5$, we find that

$$T(P_{LU}) = t_l(1, 1) + N(4t_n + t_2 + t_3 + t_5) - t_2 + t_4 - t_5 + t_n, \qquad (1.2.4.34)$$

where $t_l(1, 1)$ is the time of termination of loading program $P_{LU}^{\mathrm{loc}}(n)$ into processor (1, 1) by an optimal loader.

1.2.4.3 A Parallel Program for Solving a System of Equations $Ax = b$

Statement of the problem. Given a homogeneous multiprocessor computational system, specify a parallel program for solving a system of linear equatons $Ax = b$ $(x, b \in R^N$, $A = \|a_{ij}\|_{N\times N})$ by the LU-expansion method (see Section 1.2.4.2) and estimate the time of execution of such a program.

The calculation scheme is based on consecutively solving the system of equations $Ly = b$ and $Ux = y$, where L and U are the matrices that figure in the LU-expansion of matrix A. Here is the text of a possible local program for solving the problem. We assume that the matrices L and U are stored in the memory of the elementary processors, vector b is stored in the left memory modules, and vector x will, after the program is executed, be stored in the upper memory modules (see Figure 1.5).

Program $P^{\text{loc}}_{Ax=b}$ (n):

$$
\begin{aligned}
&\text{IF } ŏ\,1 < ŏ\,2 \Rightarrow \text{GET } 1,\ Y;\ \text{GET } 2,\ R;\ \text{PUT } 4,\ R;\\
&\qquad Y = Y - L * R;\ \text{PUT } 3,\ Y\\
&\# ŏ\,1 = ŏ\,2 \Rightarrow \text{GET } 1,\ Y;\ \text{PUT } 3,\ Y;\ \text{PUT } 4,\ Y;\\
&\qquad \text{GET } 3,\ Y;\ X = Y/U;\ \text{PUT } 2,\ X\\
&\# ŏ\,1 > ŏ\,2 \Rightarrow \text{GET } 1,\ Y;\ \text{PUT } 3,\ Y;\ \text{GET } 4,\ X;\\
&\qquad \text{PUT } 2,\ X;\ \text{GET } 3,\ Y;\ Y = Y - U * X;\ \text{PUT } 1,\ Y\\
&\text{FI.}
\end{aligned}
\tag{1.2.4.35}
$$

The corresponding Bellman equation for a nonhomogeneous medium $M^{P_{Ax=b}}$ with base $\mathscr{B}$ (n), whose power is $|K(n)| = 7 + \operatorname{sgn}(n_1 - n_2)$ can be reduced to a system of equations for homogeneous computational media $M^{(\alpha)}$, $\alpha = 1, 2, 3$, whose lattices $\Omega^{(\alpha)}$ are depicted in Figure 1.7:

$$\Omega^{(1)} = \{n \mid n_1 < n_2\},$$
$$\Omega^{(2)} = \{n \mid n_1 = n_2\},$$
$$\Omega^{(3)} = \{n \mid n_1 > n_2\}.$$

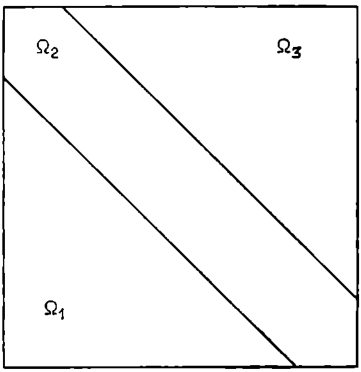


Fig. 1.7

We will not carry out the actual calculations, since they are similar to those carried out in Sections 1.2.1 and 1.2.2. The final result is as follows. The time $T(P_{Ax=b})$ of termination of program $P_{Ax=b}$, provided that program $P^{\text{loc}}_{Ax=b}(n)$ is loaded by an asymptotically optimal loader (see Section 1.2.5 below), is equal to

$$T(P_{Ax=b}) = t_l(1, 1) + N(8t_n + 2t_2 + t_1) - 2t_2 - 3t_n + t_{\leqslant},$$

where $t_l(1, 1)$ is the time of termination of the loading of program $P^{\text{loc}}_{Ax=b}$ into processor $(1, 1)$, $t_1 = t_/ + t_=$, and $t_2 = t_* + t_- + t_=$.

1.2.5 The Design of an Asymptotically Optimal Loader for a Homogeneous Computational System

In this section we give a program that realizes, in the minimum possible time, the loading of local programs into the processors of a homogeneous computational system when the number of such processors grows without limit.

It follows from the formulas for solving the problems discussed above that the time $T(P)$ of execution of a parallel program $P = \{P^{\mathrm{loc}}(n),\ n \in \Omega\}$ depends on the time $t = t_{\mathrm{l}}(P^{\mathrm{loc}}(n))$ required to load the local program into processor n; this means that the loading time is a parameter in problem (1.2.4.2); hence, with respect to this parameter we can optimize $T(P)$.

The process of loading the local programs $P^{\mathrm{loc}}(n)$ is carried out via a special parallel program $P_{\mathrm{l}} = P_{\mathrm{l}}^{\mathrm{loc}}(n)$, $n \in \Omega$, the loader of the computational system, stored in the ROM of each processor n. It has been found that as the dimensionality of the problem increases, or as $n \to \infty$, the steady-state Bellman equation applied to a special discrete computational medium can be used to specify an asymptotically optimal loader, that is, a program P_{l}^{*} for which the execution time is minimal with $n \to \infty$.

Let us detail the statement of the problem. By $\mathcal{P}_{\mathrm{l}}(\mathrm{Ar})$ we denote the set of all loading programs for a given architecture of a homogeneous computational system, and by $T(P_{\mathrm{l}})$ we denote the time of execution of a parallel program $P_{\mathrm{l}} \in \mathcal{P}_{\mathrm{l}}(\mathrm{Ar})$ on processor n, that is,

$$T(P_{\mathrm{l}}) = t(P_{\mathrm{l}}^{\mathrm{loc}}(n)) - t_0,$$

where t_0 is the start time of the program P_{l}, t_0 is a constant independent of n, and $t(P_{\mathrm{l}}^{\mathrm{loc}}(n))$ is the termination time of the operation of the program of loading into processor n.

Let us consider the problem of choosing the optimal loading program P_{l}^{*} (the optimal loader) for a given homogeneous computational system:

$$T(P_{\mathrm{l}}) \to \min \ \forall P_{\mathrm{l}} \in \mathcal{P}_{\mathrm{l}}(\mathrm{Ar}).$$

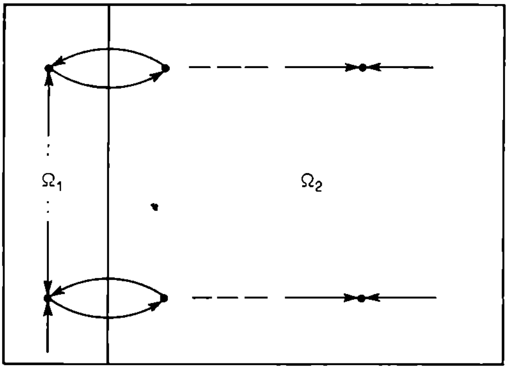
For the sake of definiteness we will consider in what follows a model of a homogeneous computational system with a two-dimensional lattice $\Omega = \{n = (n_1, n_2),\ 1 \leqslant n_i \leqslant N,\ i = 1, 2\}$. Since the time of loading into processor n is a linear function of the length ζ of the program $P^{\mathrm{loc}}(n)$ being loaded, where $\zeta = |P^{\mathrm{loc}}(n)|$, it is natural to assume that $t(P_{\mathrm{l}}^{\mathrm{loc}}(n))$ does not exceed $\zeta \sum_{i=1}^{2} C_i^1 n_i = \zeta \langle C^1, n \rangle$, $C_i^1 \geqslant 0$, $i = 1, 2$. Let

$$t(P_{\mathrm{l}}^{\mathrm{loc}}(n)) = \zeta \langle C^1, n \rangle + \langle C^2, n \rangle + C_3 \zeta + C_4, \tag{1.2.5.1}$$

where $C^2 = (C_1^2, C_2^2)$, $C_i^2 \geqslant 0$, and C_3 and C_4 are constants determined by the times of execution of the various elementary operations that enter into the trace of program P_l, $C_i^k = C_i^k(P_l)$, $k = 1, 2$. With functions of the type (1.2.5.1), the time of execution of program P_l on processor $n = (n_1, n_2)$ becomes

$$T(P_l) = \zeta \langle C^1, n\rangle + \langle C^2, n\rangle + C_3\zeta + C_4 - t_0. \qquad (1.2.5.2)$$

Hence, the objective function (or criterion function) of problem (1.2.5.2) becomes asymptotically optimal as $n \to \infty$ if there exists



$\Omega_1 = \{n \mid n_1 = 1\}$

$\Omega_2 = \{n \mid n_1 > 1\}$

Fig. 1.8

a program P_l^* for which the following holds true:

$$\forall P_l \in \mathscr{P}_l(\mathrm{Ar}) \quad C_i^1(P_l) \geqslant C_i^1(P_l^*) = 0, \quad C_i^2(P_l) \geqslant C_i^2(P_l^*), \quad i = 1, 2. \qquad (1.2.5.3)$$

A program P_l^* satisfying (1.2.5.3) is said to be the optimal loader for a computational system. Here is the text of such a program.

Program $P_l^{*\mathrm{loc}}(n)$:

```
S = ⟨NUL⟩; ADR = ⟨ADR₀⟩;
DO S ¬= ⟨EOF⟩ ⇒
   IF ω 1 = 1 ⇒
    GET 3, S; PUT 4, S;
    # ω 1 > 1 ⇒
      GET 1, S;
```

FI;
PUT 2, S;
ADR = ADR + 1;
MEMORY (ADR) = S;
OD.

Here ⟨NUL⟩ is the nil symbol, ⟨EOF⟩ the end-of-file (or program) symbol, and ⟨ADR_0⟩ is the address from which the program is loaded into the processor memory, MEMORY.

The time of execution $T(P_1^*)$ of the above program is determined by solution of the appropriate Bellman equation in a discrete medium (Figure 1.8). As a result of calculations similar to those carried out in Section 1.2.4.2, this equation can be reduced to a one-dimensional scalar equation of the type similar to that discussed in the example in Section 1.2.2.4. The final result is

$$T(P_1^*) = t_1 + (t_2 + 3t_p)\,\zeta + (n_1 + n_2 - 2)\,t_p - t_0.$$

Here t_1 is the total time spent on execution of the operators S = ⟨NUL⟩; and ADR = ⟨ADR_0⟩; and on verification of the relationship $S \urcorner =$ ⟨EOF⟩, t_2 is the total time of execution of the operators ADR = ADR + 1; and MEMORY (ADR) = S; and of verification of the relationship $S \urcorner =$ ⟨EOF⟩.

1.3 Optimization Problems of Functioning of Computational Systems

In this chapter we will discuss the main optimization problem that emerges in the functioning of operating systems (OS) of computational media, that is, the optimal organization of the execution of the users' instructions.

1.3.1 Effectiveness Criteria of the Functioning of Computational Systems

This section will formulate the general requirements of the optimization of systems of parallel data processing and give a general formulation of the problem.

An operating system is usually understood to be software that realizes the interface between the users' instructions (or programs) and the resources of a computational system. The principles of design of operating systems have been discussed in great detail in the works of Soviet and other authors [1.18, 1.25-1.28].

Without assuming a formal approach, we may say that the problem lies in allocating the resources to the set of users' instructions so as to optimize a measure of effectiveness of the execution of the in-

structions, say, the average execution time or the maximal execution time. The choice of measure of effectiveness is determined by two trends in the organization of parallel data processing: (1) the attainment of high total productivity in executing various usually small and weakly interacting instructions and (2) the attainment of maximal productivity in solving a single large problem [1.17, 1.18].

Here is a formal statement of the problem. Suppose that we have specified a set $P = \{p_1, p_2, \ldots, p_m\}$ whose elements are the resources of the computational medium. A system of instructions is a pair (X, ρ), where $X = \{x_1, x_2, \ldots, x_n\}$ is a set whose elements are the instructions x_j, $j = 1, \ldots, n$, and ρ is the partial ordering relation on X.

Remark 1.3.1.1. For the sake of definiteness the resources of a computational medium are understood in what follows to be the processors of that computational medium. However, the discussion can be generalized to other types of resources. In homogeneous computational media the processors are identical both in functional capability and in speed. The partial ordering relation ρ on X imposes restrictions on the sequence in which the instructions are executed: $x_i \rho x_j \Leftrightarrow$ means that instruction x_i precedes instruction x_j if x_i must be completed before x_j. In this section we will call the relation ρ the instruction-precedence relation.

Let us denote by $\tau_{ij} > 0$ the time of execution of instruction x_j, $j = 1, \ldots, n$, on processor P_i, $i = 1, \ldots, m$. For a homogeneous computational medium, $\tau_{ij} = \tau(x_j)$ is the time of execution of instruction x_j on any processor. By W_i, $i = 1, \ldots, n$, we denote the specific cost of stay of instruction x_i in the computational medium.

Remark 1.3.1.2 The cost of stay of an instruction in a computational medium depends on the parameters of the instruction. These may be the memory size (the number of cells) required for storing the instruction, for instance, or the reaction time of the system when the instruction is executed in real time.

The problem of organizing the execution of instructions in a computational medium consists in building a schedule s, a pair of mappings $f\colon X \to R^1_+$, $g\colon X \to P$ $(s \overset{\text{def}}{=} (f, g))$ for which the following properties hold true:

$$(1)\ (\forall x_i, x_j \in X,\ g(x_i) = g(x_j) = k,\ k \in 1, \ldots, n)$$

$$\Rightarrow \chi_{f(x_i),\, f(x_i)+\tau_{k_i}}(t) \cdot \chi_{f(x_j),\, f(x_j)+\tau_{k_j}}(t) = 0,$$

where $\chi_{a,b}(t) = 1 - \theta(t - b) - \theta(-t + a)$ is the characteristic function of segment $[a, b]$, with $\theta(t)$ being Heaviside's function; and

$$(2)\ x_i \rho x_j \Rightarrow f(x_i) + \tau_{g(x_i)i} \leqslant f(x_j).$$

Remark 1.3.1.3 The value $f(x_i)$ of function f at point x_i is the beginning time of execution of instruction x_i, while $g(x_i)$ defines

the processor ensuring the execution of the instruction. The first property means that each processor P_i at a fixed moment in time t can execute only a single instruction. The second property realizes in time the partial ordering relation ρ on the set of instructions X.

The problem of optimal organization of execution of a system of instructions consists in building a schedule s^* that delivers an extremum to one of the following measures of effectiveness of operation of a computational medium:

$\mu_1 = \min\limits_s \max\limits_{1\leq i\leq n} f(x_i)$, the minimal possible time of termination of all instructions;

$\mu_2 = \min\limits_s n^{-1} \sum\limits_{i=1}^{n} W_i f(x_i)$, the smallest cost-weighted mean of the time during which the executed instructions are kept in the computational medium;

$\mu_3 = \min\limits_s n^{-1} \sum\limits_{i=1}^{n} (f(x_i) - \tau_{g(x_i)i}) W_i$, the minimal cost-weighted mean of the time during which the instructions wait to be executed in the computational medium;

$\mu_4 = \min\limits_s n^{-1} \sum\limits_{i=1}^{n} (f(x_i) - d_i)_+ W_i$, the minimal cost-weighted-mean of delay of execution of the instructions in relation to the given directive times d_i, $i = 1, \ldots, n$, in which the instructions were supposed to be executed (here we employ the notation $\Phi(x)_+ = \max\{0, \Phi(x)\}$); and

$\mu_5 = \min\limits_s \lambda^{-1}(s) \sum\limits_{i=0}^{n-1} (n-i)(f(x_{p_{i+1}}) - (f(x_{p_i}))$, with $f(x_0) = 0$

and $\lambda(s) = \max\limits_{1\leq i\leq n} f(x_i)$, the quantity characterizing the minimal expected demand of the system of instructions in processors. It is assumed that $f(x_{p_{i+1}}) \geqslant f(x_{p_i})$, $i = 0, \ldots, n-1$.

Remark 1.3.1.4 For a homogeneous computational medium, criterion μ_3 is equivalent to μ_2, while criterion μ_4 makes it possible to pose the problem of minimizing the number of processors that can ensure the general directive time limit $d = d_i$, $i = 1, \ldots, n$, for execution of the instructions.

Passage to the limit in the small parameters $h_1 = 1/m$ and $h_2 = 1/n$ as $m \to \infty$ and $n \to \infty$ in discrete computational media yields measures of effectiveness of the execution of a system of instructions in continuous computational media. For example, if we send n to ∞

in μ_5, we get $\mu_5 = \min_s \lambda^{-1}(s) \int_{\oplus}^{\lambda(s)} N(t)\,dt$, where $N(t)$ is a smooth monotonic decreasing function, the number of instructions not executed by time t.

Assertion 1.3.1.1 *All effectiveness criteria mentioned above are linear in spaces with values in one of two semi-rings,*

$$A_1 = (R^1_+ \cup \{+\infty\},\ \oplus = \min,\ \odot = \max),$$
$$A_2 = (R^1_+ \cup \{+\infty\},\ \oplus = \min,\ \odot = +).$$

As concrete examples we will formulate and solve three problems of optimal control of the resources in homogeneous computational media. The first two deal with the optimization in the given criteria of parallel data processing in homogeneous computational media, while the third deals with the control of external switching.

1.3.2 Optimal Organization of Parallel Data Processing

In this section we give algorithms for the solution of two main problems emerging in optimal parallel calculations.

The main difficulty in parallel calculations lies in the need to maintain the partial order and temporal sequence in carrying out the separate computational instructions or assignments that are worked into the algorithms or programs as a result of ordered data processing [1.18]. The parallelism of calculations is carried out between the individual processors that constitute the computational system. Here two main problems emerge, in a sense reciprocal:

(a) given an algorithm and a fixed time interval assigned for execution of the algorithm, find the minimal number of processors constituting a homogeneous computational system; and

(b) given an algorithm, find the minimal time of its execution by a given homogeneous computational system that incorporates m processors united by a common main storage.

Essentially these problems can be reduced to the problem of optimal distribution of a set of data-correlated operators or, in other words, of a partially ordered set of the calculation assignments distributed between the processors. Here is a formal statement of this problem.

Suppose that realizing an algorithm on a homogeneous CS with a general main storage (or memory) requires realizing a system of instructions (X, ρ). Here $X = \{x_1, x_2, \ldots, x_n\}$ may be, for instance, a set of problems, operators, or commands, while the ordering relation ρ is determined by the information correlations between the instructions. Instruction x_i requires $\tau(x_i)$ units of time for its execution and may be carried out by any processor in the system.

Problem 1. It is required to assign the instructions to the processors in such a way that (1) the time of execution of all instructions assigned to each processor must not exceed a given time T, (2) the instruction-precedence relation ρ (which is determined by the data correlations between the instructions) must be satisfied, and (3) the number of processors must be minimal.

Problem 2. It is required to assign the instructions to the processors in such a manner that (1) the number of processors m is given, (2) the instruction-precedence relation (which is determined by the data correlations between the instructions) must be satisfied, and (3) the time $T = \max\limits_{1 \leqslant i \leqslant m} t_i$, where t_i is the time of execution of all instructions on processor i, must be minimal.

The requirement that the instruction-precedence relation be satisfied can be expressed in the form of a cyclic graph $G_\rho = (X, \Gamma_\rho)$, which represents the instruction-precedence relation ρ. The vertices x_i and x_j of graph G_ρ are connected by the arc $(x_i, x_j) \in \Gamma_\rho$ if instruction x_i must precede instruction x_j. We will say that a subset of vertices $Q \subseteq X$ of graph G_ρ is allowed if there is no arc $(x_i, x_j) \in \Gamma_\rho$ that leads from $X \setminus Q$ to Q $(x_i \in X \setminus Q,\ x_j \in Q)$. The predicate defined in this manner on the set 2^X will be denoted by $\mathscr{D}$.

The solution to Problem 1 can be expressed in terms of the solution to the appropriate generalized Bellman equation in the space of functions with values in the Abelian semi-group $R_\oplus = (R, \oplus)$ (see Section 1.4) for a discrete medium uniquely determined by the statement of the initial problem, while the coefficients of the equation, which are endomorphisms of semi-group $R_\oplus$, are fixed by the temporal parameters $\tau(x_i)$ $(x_i \in X)$ and T, $T > 0$.

Let us consider the discrete medium $M = (\widetilde{X}, \widetilde{\Gamma}, H, R)$ with

(1) $\widetilde{X} = \{x_i \in 2^X,\ \mathscr{D}(\widetilde{x}_i)\}$, $i = 1, \ldots, k$, $k = |\widetilde{X}| \leqslant 2^n$, $\widetilde{x}_1 = \varnothing$, $\widetilde{x}_k = X$;

(2) $\widetilde{\Gamma} = \bigcup\limits_{i=1}^{n} \Gamma_i$, $\Gamma_i = \{(\widetilde{x}, \widetilde{y}),\ \widetilde{y} \setminus \widetilde{x} = \{x_i\},\ \widetilde{x}, \widetilde{y} \in \widetilde{X}\}$;

(3) $R_\oplus = (R, \oplus)$, $R = \{1, 2, \ldots, n\} \times \{0, T+1\}$, and the semi-group operation $\oplus$ is determined by the relation

$$a \oplus b = \begin{cases} (a_1, a_2), & (a_1 < b_1) \vee (a_1 = b_1 \wedge a_2 \leqslant b_2), \\ (b_1, b_2), & (a_1 > b_1) \vee (a_1 = b_1 \wedge a_2 > b_2), \end{cases}$$

where $a = (a_1, a_2)$, $b = (b_1, b_2) \in R$. Obviously, the neutral element has the form $\varnothing_R = (n, T+1)$. Endomorphism H is defined in the following manner:

$$H(u) = \begin{cases} I(u) & \text{if } u \in \widetilde{\Gamma}, \\ \varnothing & \text{if } u \notin \widetilde{\Gamma}, \end{cases}$$

where

$$\forall u \in \Gamma_i,\ L(u)(a) = \begin{cases} (a_1,\ a_2+\tau(x_i)) & \text{if } a_2+\tau(x_i) < T, \\ (a_1+1,\ \tau(x_i)) & \text{if } a_2+\tau(x_i) > T, \\ a_1 < n,\ i=1,\ \ldots,\ n, & \\ \mathbb{O}_R \text{ if } a_2+\tau(x_i) > T,\ a_1 = n. & \end{cases} \tag{1.3.2.1}$$

Let us describe the algorithm for solving Problem 1:

(1) put $\mathscr{F}(\tilde{x}) = \begin{cases} (1,\ 0) & \text{if } \tilde{x} = \tilde{x}_1, \\ \mathbb{O}_R & \text{if } \tilde{x} \neq \tilde{x}_1; \end{cases}$

(2) apply the H-scheme (see Section 1.5.2) to solve the Bellman equation for the discrete medium M:

$$s = Hs \oplus \mathscr{F}, \tag{1.3.2.2}$$

where $\mathscr{F}$ is defined in item (1);

(3) reconstruct the route μ_{opt} from solution s (see Section 1.5.3);

(4) terminate operation; the result is as follows: instruction x_i is executed on processor P_j with number $j = a_1$ over the time interval $[a_2 - \tau(x_i),\ a_2]$, where $(a_1,\ a_2) = s(\tilde{x})$. Here s is the solution to Eq. (1.3.2.2) at point $\tilde{x}$ satisfying the condition $\Gamma_i \cap \mu_{\text{opt}} = \{(\tilde{y},\ \tilde{x})\}$.

Here is the algorithm for solving Problem 2:

(1) put $T_1 = \max\limits_{x_i \in X} \tau(x_i) - 1$, $T_2 = \sum\limits_{i=1}^{n} \tau(x_i)$, and $T = T_2$;

(2) apply the algorithm for solving Problem 1;

(3) suppose that $s(\tilde{x}_k) = (a_1,\ a_2)$: if $m < a_1$, then put $T_1 = T$, otherwise put $T_2 = T$;

(4) if $T_2 - T_1 \leqslant \varepsilon$, where ε is the required accuracy of solution, then go on to item (5), otherwise put $T = (T_1 + T_2)/2$ and go over to item (2);

(5) terminate operation; the result is the route μ found in item (2) of the algorithm at $T = T_2$. (If $\tau(x_i) \in \mathbb{Z}_+$ and $\varepsilon = 1$, the solution is exact.)

1.3.3 Optimal Control of Switching

Here we will discuss the organization of switching in a computational medium. We will demonstrate that the problem of optimal control of external switching on the basis of ω-networks with determinate servicing can be reduced to solving the Bellman equation with semi-group operations $\oplus = \min$ and $\odot = \max$.

The effectiveness of a computational medium depends essentially on the organization of switching in the medium [1.17, 1.18]. Two types of switching are distinguished: internal and external. Internal switching is the program-controlled variation of connections between the elementary processors of the medium. External switching is the program-controlled variation of connections between the computational medium and peripheral devices (external memory, terminals, data stations, and the like).

It is natural to understand internal switching control in a computational medium as a transformation of the coefficients in the generalized Bellman equation describing the propagation of local data alterations in the computational system (see Section 1.2.3). Then the problem of optimal control of internal switching can be reduced to such a choice of the coefficients of the equation that the solution to the equation will guarantee an extremal value of one of the objective functions of the effectiveness of CS functioning. For instance, the optimal value of the activity range of multiprocessor homogeneous CS as a function of the architecture parameters of the CS can be found by solving the Bellman equations constructed in Section 1.2.2. Other possible criteria of effectiveness of computational systems were discussed in Section 1.3.1. Two main methods of organizing external switching are known, spatial and temporal, but combinations of the two are also possible [1.16-1.18].

An example of temporal switching is the so-called switching with a common bus [1.18]. In this case all devices subject to switching are connected to a single channel, which makes it possible to connect, at each moment in time, only one peripheral device with a given elementary processor of the medium. In spite of the low cost and simplicity of realization of this type of switching, it has significant deficiencies: a low throughput and high delays in data transfer.

An example of spatial switching is the array switching [1.18], which ensures a high throughput of data transfer thanks to the possibility at each moment in time of pair switching of any peripheral device with an elementary processor of the medium.

Specialists now agree that the most effective and economical way of realizing external switching on VLSI schemes is to employ the so-called ω-network [1.16, 1.18], which combines the merits of spatial and temporal switching. For one, the complexity of realization of a ω-network is proportional to the logarithm of the number of devices included in the switching process [1.18].

The problem of optimal control of an ω-network (with determinate servicing). The model of operation of an ω-network. The structure of an ω-network is as follows (Figure 1.9): an ω-network contains (1) 2^{n+1} elementary switches $X_{\text{in}} = \{x_i,\ i = 0, \ldots, 2^n - 1\}$ taking part in the switching process, 2^n peripheral devices at the input port of the ω-network, and 2^n elementary processors $X_{\text{out}} = \{y_i,\ i =$

0, . . ., 2^{n-1}} and (2) $2^{n-1} \times n$ elementary switches, each of which may be in one of two states and ensures the commutation (steady-state connection or link) of two of its inputs with two of its outputs (see Figure 1.9).

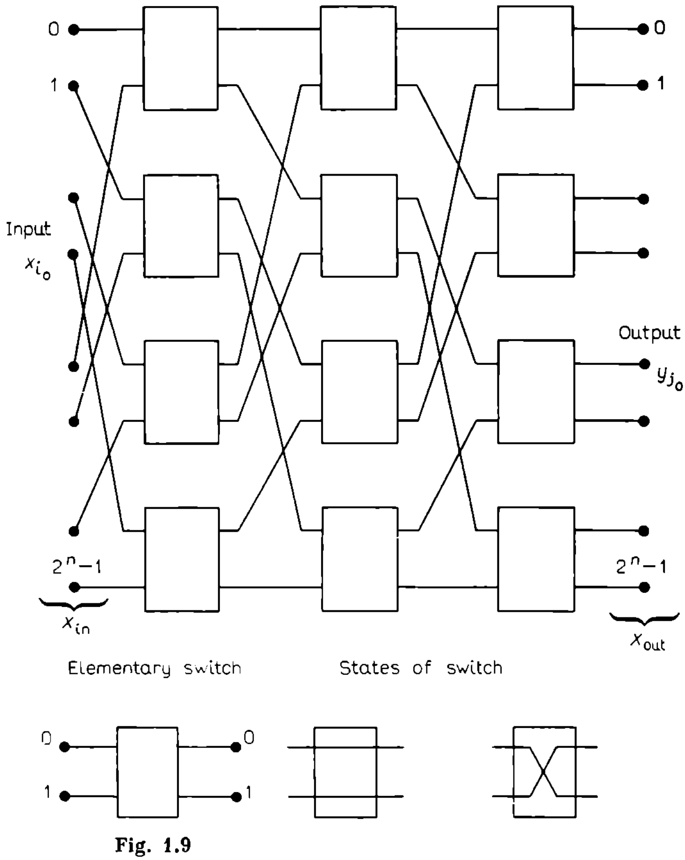


Fig. 1.9

The steady state of an ω-network is given by the collection of the states of all of the network's elementary switches. The input port of an ω-network is used to receive the instructions for data processing in the computational medium from a peripheral device. A special processor, which controls the operation of the ω-network, must determine ("commutate") the elementary processor that will receive the

instruction for data processing. In other words, this processor must connect the input port of the ω-network (specifically, the peripheral device $x_i \in X_{\text{in}}$) with the output port (specifically, with one of the processors of the computational system), which means that it must find a route for transferring the instruction through the switches of the ω-network (i.e. through the "free" connections between the switches) that leads from $x_i \in X_{\text{in}}$ to the set X_{out}.

At each moment in time t, the set of all possible routes of data transfer according to the instruction at the input port of the ω-network depends on the steady state of the ω-network at time t and on the instructions that were at the input port of the ω-network by time t (a fraction of the switches and steady-state connections between the switches are engaged in carrying out these instructions). Let us define the dynamical state $C_\omega(t, x_0(t))$ of an ω-network as the set of all routes of data transfer admissible at time t that lead from $x_0(t) \in X_{\text{in}}$ into X_{out} (routes that do not intersect through the switches). Let us assume that the operation of the ω-network is determined, that is, for each instruction received by the ω-network we know the time of its processing and the dynamical state $C_\omega(\tau, x_0(\tau))$ of the ω-network at $\tau \in [t, T_0]$, $T_0 > t$, where T_0 is the time of termination of the operation of the ω-network involved in processing all the instructions received by time t.

Statement of the problem. Suppose that at time t the input port of an ω-network receives an order for a peripheral device $x_0 \in X_{\text{in}}$. This order must be fulfilled in a minimal possible time with due regard for the restrictions imposed on the dynamical state of the network at $\tau \in [t, T_0]$ by the system of orders received at the input port of the ω-network by time t.

To solve the formulated problem, we give the appropriate Bellman equations in the space of functions with values in the semi-ring $A = (R^1 \cup \{\pm\infty\},\ \oplus = \min,\ \oplus = \max)$ for a discrete medium whose architecture is depicted in Figure 1.10. This equation has the form

$$s_{k+1}(i, j) = \min_{\substack{0 \leqslant i' \leqslant 2^n - 1 \\ 0 \leqslant j' \leqslant n}} \max\{s_k(i', j'), a(i, j, i', j')\}$$

$$k = 0, 1, 2, \ldots, \tag{1.3.3.1}$$

$$s_0(i, j) = \begin{cases} \infty & j \neq 0, \quad i \neq i_0, \\ t, & j = 0, \quad i = i_0. \end{cases} \tag{1.3.3.2}$$

Here i_0 is the number of a given peripheral device $x_0 \in X_{\text{in}}$, (i, j) are the coordinates of a steady-state connection in the system, i is the number of a vacant output of an elementary switch in the ω-network positioned in the layer with number j (for $j > 0$) or a peripheral device (at $j = 0$), $s_k(i, j)$ is the moment in time when there is a free route of length k (i.e. consisting of k steady-state connections)

leading from the peripheral device x_{i_0} specified in the order to the elementary switch with output i in the jth layer. The coefficients in Eq. (1.3.3.1) are uniquely determined by the given dynamical states $C_\omega(\tau, x_0(\tau))$, $\tau \in [t, T_0]$:

$$a(i, j, i', j') = \begin{cases} t_{ij} \text{ if } j - j' = 1,\ i = \varepsilon + 2i' (\mathrm{mod}\ 2^n),\ \varepsilon = 0, 1, \\ \infty \text{ otherwise} \end{cases} \tag{1.3.3.3}$$

where t_{ij} is the time when the switch with number $([i/2], j)$ is free from fulfilling previous orders (here [ ] denotes the integral part of a number).

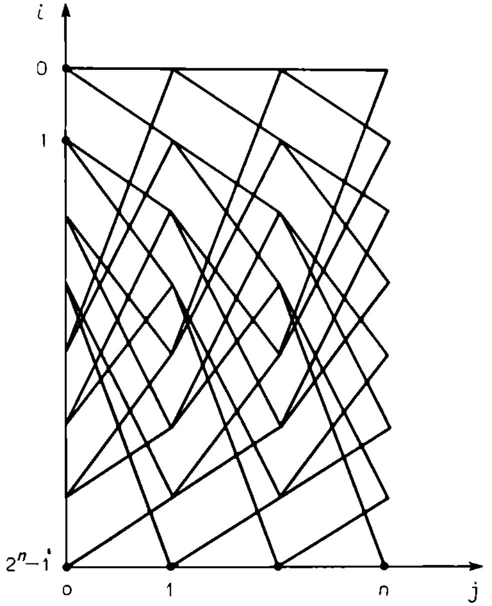


Fig. 1.10

Obviously, the solution to problem (1.3.3.1), (1.3.3.2) at $k = n$ determines the optimal (i.e. the shortest in time) way of connecting the peripheral device $x_0 \in X_{\text{in}}$ with the output port of the ω-network. The number of the i^*th processor in the system (i^*, n) that receives the order from x_{i_0} to process the information can be found from the equation $i^* = \arg \min\limits_{0 \leqslant i \leqslant 2^n - 1} s_n(i, n)$, and the sought route $\mu^* =$

$\mu(i_0, 0) \to (i^*, n))$ can be reconstructed uniquely from the found number i^* as follows: if $i^* = (k_1, k_2, \ldots, k_n)$ is the binary representation of number i^*, the sequence $(i_0, 0)$, $(2i_0 \pmod{2^n} + k_1 = i_1, 1)$, $(2i_1 \pmod{2^n} + k_2, 2)$, . . ., $(2i_{n-1} \pmod{2^n} + k_n, n)$ specifies the vertices of the optimal route μ^*. The method of reconstructing the route μ^* and building the solution to the Bellman equation (1.3.3.1) follows from the algorithmic procedure (the H-scheme) discussed in Section 1.5.2 (see also [1.24]).

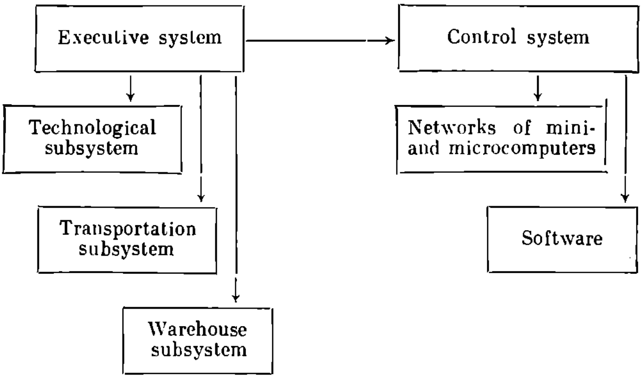
1.4 Flexible Automatic Manufacturing of Computational Media

As yet there is no flexible automatic manufacturing of computers. However, in long-term projects for designing fifth generation computers the organization of flexible automatic manufacturing systems has been discussed [1.29-1.31]. There is therefore a need to study related mathematical problems. First, there are the problems pertaining to all flexible automatic manufacturing and, for one, the flexible automatic manufacturing of many "small" components, or continuous computational media, say, the flexible automatic manufacturing of the element base of fifth generation computers using VLSI, in which mathematical models are employed for manufacturing thin films, integrated circuits, and integral-optics elements. In what follows we consider the mathematical problems involved in flexible automatic manufacturing from this point of view bearing in mind the large parameter that appears in the problems arising in this connection.

The definition of flexible automatic manufacturing. At present there is no universally accepted definition of flexible automatic manufacturing. We will adhere to the following, most widely accepted, formulation of this concept. Flexible automatic manufacturing is a manufacturing unit—production line, bay, shop, or plant—that functions on the basis of unmanned technology, program control, and group organization of manufacturing. Structurally a flexible automatic manufacturing system consists of an executive system that incorporates technological, transportation, and warehouse subsystems, and a control system that coordinates the functioning of the executive subsystems. The control system consists of mini- and/or microcomputers united by data transmission links, and software that controls both the separate equipment units and the system as a whole (see Table 1.4.2). Program control with computers equipment and manufacturing processes ensures the flexibility of the automatic manufacturing process, that is rearrangement of all the executive components in the manufacturing system when the objects of manufacture are changed. The group organization of manufacturing

is based on special-purpose specialization of the bays and shops and a unified group form of organization of the manufacturing processes [1.32, 1.33].

Table 1.4.2 The Structure of Flexible Automatic Manufacturing Systems



1.4.1 Classification of the Mathematical Problems Associated with Flexible Automatic Manufacturing

In this section we suggest a new classification of the mathematical problems associated with the design and functioning of flexible automatic manufacturing systems. The discussion is based on the theory put forward in the previous sections.

There now exists a vast literature on work connected with flexible automatic manufacturing. However, a large portion is devoted to the manufacture-organizational principles of designing flexible automatic manufacturing. Practically no publications discuss the mathematical aspects of such design and the functioning of such systems.

We distinguish six large mathematical problems in flexible automatic manufacturing. We believe that all mathematical problems arising in flexible automatic manufacturing at the stage of design of such a system and its functioning, except for problems of program control of production equipment, can be treated as linear problems in spaces with values in an abstract semi-ring, which means that the generalized Bellman equation can be used to solve such problems, both in the discrete variant (see Section 1.1.10) and in the continuous case (see Section 1.0), when in stating a problem we can specify a small parameter in which passage to the limit is possible. For instance, in planning problems the quantity reciprocal to the number of performed operations can serve as such a parameter.

The first problem in designing a flexible automatic manufacturing system, a problem inherent in any such system, is the classification of the manufacture products by their structural-production characteristics with the aim of their further grouping. This problem can be solved by applying methods of cluster analysis [1.34-1.36], which employ concrete realizations of the abstract semi-ring, these realizations are determined in the ways in which the distance between objects in the classification space is specified.

We now wish to state the classification problem. Let $(\mathcal{L}, d)$ be the set of objects $\mathcal{L} = \{x, y, z, \ldots\}$, $|\mathcal{L}| = n$, with a specified generalized "metric" d: $\mathcal{L} \times \mathcal{L} \to R$. We wish to find the partition J^* of set $\mathcal{L}$ into a given number m of groups of objects $\{J_i^*, i = 1, \ldots, m\}$ that minimizes the generalized sum of the distances between the objects within each group:

$$J^* = \arg \min_{J \in \mathbb{J}} \bigodot_{i=1}^{m} \bigodot_{x,\, y \in J_i} d(x, y), \tag{1.4.1.1}$$

where $\mathbb{J}$ is the set of all possible partitions $J = \{J_1, J_2, \ldots, J_m\}$ of the set $\mathcal{L}$ and $\odot$ is the generalized multiplication sign (see Section 1.1.1).

Problem (1.4.1.1) is linear in the space of functions with values in the semi-ring $A = (R, \oplus = \min, \odot)$. In concrete classification problem the generalized metric on the set of objects is determined by the numerical values of the characteristics, that is, the coordinates $(x_1, x_2, \ldots, x_n)$ of object $x \in \mathcal{L}$ in the characteristics space R^n. For instance, the metric $d(x, y)$ can be specified by one of the following formulas:

$$d(x, y) = \max_{1 \leqslant i \leqslant n} |x_i - y_i| \quad \text{(Manhattan distance [1.37])},$$

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i| \quad \text{(Hamming's distance [1.38])}.$$

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad \text{(Euclidean distance)}.$$

For the above cases the solution to problem (1.4.1.1) is given in [1.35, 1.39].

The classification problem remains linear even when metric d is induced by a fuzzy resemblance relation τ (see [1.40, 1.41], τ: $\mathcal{L} \times \mathcal{L} \to [0, 1]$ is a reflexive ($\forall x \in \mathcal{L}, \tau(x, x) = 1$) and symmetric ($\forall x, y \in \mathcal{L}, \tau(x, y) = \tau(y, x)$) mapping. The transitive closure $\hat{\tau}$ of the resemblance relation τ is a similarity relation or a fuzzy equivalence relation (see [1.38]) possessing the properties of reflexivity, symmetry, and transitivity: $\forall x, y, z \in \mathcal{L}$, $\hat{\tau}(x, y) \geqslant \min(\hat{\tau}(x, z),$

$\hat{\tau}(z, y)$). We define a "minimax" metric d on $\mathcal{L}$ by the formula $d(x, y) = 1 - \hat{\tau}(x, y)$. The problem with the "fuzzy" metric d is linear in the space of functions with values in the semi-ring $A = (R = [0, 1], \oplus = \min, \odot = \max)$. Its solution is determined in terms of the solution of the system of Bellman equations for the shortest connecting tree [1.42]. For a large number N of graded objects it is natural to solve the classification problem via a passage to the limit in the small parameter $1/N \to 0$ in the discrete Bellman equation.

Note that the solution to the classification problem makes it possible to design a group route and operational technology of processes involving the manufacture and inspection of products [1.43].

The second problem in the design of flexible automatic manufacturing systems involves the planning of flexible automatic production lines and bays on the basis of a calculation of the optimal supply and timetables of delivery of resources and the optimization of the traffic. The mathematical statements of problems associated with this aspect are known in the literature as network flow problems [1.9, 1.37, 1.44, 1.45].

It has been established that all these problems are linear in spaces with values in certain semi-rings, say, the problem of designing the maximum flow in a stationary transport network is linear in the space with values of functions in the semi-ring $A = (R^1, \oplus = \max, \odot = \min)$. The Duhamel principle applied to the inhomogeneous Bellman equation describing the flow in a nonstationary transport network leads to a linear problem in the same space.

The problem of constructing a maximum flow in a transport network is classical. An analysis of modern results in the range of its solution shows that for arbitrary carrying capacities of the arcs the algorithms listed in Table 1.4.3, with $n = |X|$ and $m = |\Gamma|$, possess the best estimates of complexity.

Table 1.4.3

Nos	Algorithm	Temporal complexity of algorithm
1	Edmonds-Karp [1.46]	$O(nm^2)$
2	Dinits [1.47]	$O(n^2m)$
3	Karzanov [1.48]	$O(n^3)$
4	Cherkasskii [1.49]	$O(n^2m^{1/2})$
5	Galil [1.50]	$O(n^{5/3}m^{2/3})$
6	Galil-Naamad [1.51]	$O(nm \log^2 n)$
7	Sleator [1.52]	$O(nm \log n)$

In solving flow problems, algorithm designers employ various procedures for constructing optimal (in a certain sense) transport network routes that increase the magnitude of the flow. These procedures are simply different ways of solving the generalized Bellman equation (see Eq. (1.1.10.2)) in spaces with values in one of the following semi-rings: (R^1, $\oplus$ = min, $\odot$ = +), (R^1, $\oplus$ = max, $\odot$ = min), or (N, $\oplus$ = min, $\odot$ = +) (cf. [1.44]). This viewpoint makes it possible to classify the algorithms of constructing a maximal flow in terms of the coefficients of the Bellman equation or to represent the solution of the Bellman equation in the form of a "continual" integral (see formula (1.2.2.18)) defined in this case on the routes of the transport network connecting the sources with the sinks. This classification can be carried over to algorithms for solving the problem of constructing a maximal flow at a minimal cost and, hence, the transportation problem in its various modifications [1.37, 1.44].

In the case of a "saturated" and regular transport network, when the arcs in the transport network satisfy the conditions of locality and regularity (see Section 1.2.1), the number of its nodes, N, tends to infinity, and the carrying capacity is a slowly varying function on the set of arcs of the transport network, we can go over to the limit in the small parameter $1/N \to 0$ to the continuous Bellman equation and, hence, the method of characteristics and Pontryagin's maximum principle can be employed to solve this equation [1.2, 1.3].

The third problem has to deal with designing a warehouse subsystem of the flexible automatic manufacturing system so as to ensure further optimal functioning. The main mathematical problem here is the problem of optimal arrangement of the resources in the warehouse, say, by minimizing the average time for selecting the resources in store. The appropriate objective function is linear in a space of functions with values in the semi-rings considered in Section 3 in connection with solution of the problem of optimal functioning of a computational medium. Here it also proves possible to pass to the limit of a continuous Bellman equation in a natural small parameter reciprocal to the number of nomenclatures of the resources stored in the warehouse or the total resources. A number of other problems of organizational and technological scope related to the design of warehouse systems have been reported in the materials of international conferences [1.53, 1.54].

The fourth problem is that of in-line manufacture planning and control, which consists of optimal processing of information and material flows in the control and executive systems of a flexible automatic manufacturing system. The mathematical aspect of the problems arising here is the same as in Section 1.3 if the information and material flows are seen as a system of instructions and the hardware of the executive and control systems as the resources of the

computational medium necessary for executing the system of instructions. Thus, the optimization problems are linear in a space of functions with values in certain semi-rings, and there are effective algorithms (see [1.37, 1.55, 1.156]) for solving these problems.

The fifth problem is that of the functioning of the technological system of a flexible automatic manufacturing system. This has been most thoroughly studied for mechanoprocessing production. It has been found that for new flexible manufacturing systems, such as the manufacturing of computers of future generations based on the technology of fabrication of VLSI circuits and elements of integral optics and molecular electronics, the development of mathematical models of the manufacturing processes is a necessary prerequisite for solving this problem. The development of such models for processes of VLSI manufacturing is fully discussed in [1.57], while the mathematical models that have found application in integral optics are discussed in [1.58, 1.59]. The last paper in the present collection discusses the statement of the problems and the algorithms of solution that are encountered in modeling various stages in technological processes in microelectronics.

The sixth, and final, problem is that of the optimal functioning of the transportation subsystem and, at present, constitutes the main problem in the design of flexible automatic manufacturing systems, since its solution largely determines the economic effectiveness of the flexible manufacturing system as a whole.

It is interesting to see what has been written on this subject in Japan and the United States. According to statistics given in the review paper [1.60], only five percent of the time in which products are in the plant is devoted to the actual machining; the remainder is lost unproductively in transportation. Designers of flexible automatic systems (e.g. designers of production systems with industrial robots) single out the optimization of transportation operations as a special problem [1.61].

As noted by the authors of [1.62], the concept of a robotized bay is as yet only in its initial stage of development. The authors of [1.62] see the future of transportation industrial robots in the design of intelligent transportation robots that will carry the products between the bay and the principal warehouse and automatically perform materials handling. These intelligent transportation robots are beginning to be equipped with on-board microcomputers. These can "contact" a computer of a higher rank every time the robots are in the materials-handling stage and, hence, can receive instructions to change a route or a routine. The authors of [1.62] believe that there is only one step from a bay with readjustable machines controlled by a computer linked with intelligent robots and a base computer that controls the entire complex to a fully integrated production plant with minimum personnel.

At present the most common transport network functioning at the production level involves one robot. In this case the problem of the optimal planning of the trajectories of the robot displacements is reduced to the classical problem of shortest route. This problem is linear in a space of functions with values in the semi-ring $A = (R^1, \oplus = \min, \odot = +)$. Technically the transport network is realized in the form of directional induction lines built into the floor of the bay or via optical recognition of passive bands, painted or metallic [1.32, 1.63]. When the movements of the robot are organized in the "automatic-pilot" mode, the problem of calculating the optimal trajectory in a continuum arises. In this case solution to the problem is provided by Pontryagin's maximum principle for the continuous Bellman equation in a space with values in the same semi-ring.

Obviously, the effectiveness of the functioning of an intrabay transportation system grows with the number of robots that can operate simultaneously on the transport network of this system in conditions of integrated flexible automatic manufacturing. The robot's optimal movement in this case is not determined solely by fixing the initial and final positions of the robot. The optimal route must guarantee that there are no collisions with other robots. The importance of an optimal solution to this problem and the fact that the problem is not trivial since there must no collisions have been emphasized in [1.64, 1.65]. A similar problem of collisionless movements appears also in robotics in connection with the planning of optimal trajectories of separate elements of manipulators [1.65, 1.66]. Below we examine in detail this most important and as yet sunolved problem of flexible automatic manufacturing.

1.4.2 Solution of the Problem of the Optimal Functioning of the Transportation Subsystem of a Flexible Automatic Manufacturing System Serviced by Intelligent Transport Robots

We now give the mathematical statement of the problem and its reduction to three optimization problems of collisionless movements of robots. Here, as in the optimization problems linear in spaces of functions with values in an Abelian semi-ring (see Sections 1.2 and 1.3), the method of solving these three problems is based on the fact that they are linear but in a space of functions with values in an Abelian semi-group. The appropriate theory of the Bellman equation in such spaces and the numerical algorithms for solution of this equation will be developed in Section 1.5.

1.4.2.1 A Model for the Transportation Subsystem in a Flexible Automatic Manufacturing System

The model of a transport network will be represented by a symmetric directed graph $G = (X, \Gamma)$, $X = X_1 \cup X_2 \cup X_3$, $|X| = n$, where X_1 is the set of nodes at which materials handling

may be carried out, X_2 is the set of crossings in the transport network, X_3 is the set of depots, that is, the places to which robots not engaged in operations can recede, and $(x_i, x_j) \in \Gamma$ is an arc corresponding to an elementary displacement of the robot from node x_i to node x_j. On each arc (x_i, x_j) the time of motion along the arc, $\tau_{ij} > 0$, is specified.

Suppose that the transport network is serviced by N robots. The servicing amounts to fulfilling the instructions that the transportation system has received to transfer resources between an arbitrary pair of nodes belonging to X_1. Suppose that by the time T_0 that the system has received a specific instruction, some of the robots (and perhaps all) are servicing the instructions that the system received at times $t < T_0$ (a robot may service several instructions only in a series). In other words, it is assumed that at each moment $t \geqslant T_0$ the state of any one of the N robots is known, and each robot may carry out materials handling at a node belonging to X_1 or may move along an arc belonging to Γ and pass through a node belonging to X_2 or may recede into a depot (i.e. at a node belonging to X_3). Thus, for $t \geqslant T_0$, (a) at each node $x_i \in X$ there is defined a set of moments b_i at which x_i is vacant, that is, moments at which there is not a single robot at vertex x_i carrying out instructions received earlier, (b) for each arc (x_i, x_j) there is known the set of moments c_{ij} at which a robot may start moving along the arc and during the time interval τ_{ij} of motion will not collide with the other robots. It is natural to assume that each node x_i belonging to X_3 is vacant at any moment in time $(b_i = (T_0, +\infty),\ x_i \in X_3)$.

1.4.2.2 Statement of the Problem of Optimal Operational Control of Materials Handling in the Transportation Subsystem

Suppose that at time T_0 an instruction to transfer a resource from node $x_{i_0} \in X_1$ to node $x_{j_0} \in X_1$ has been received by the transportation subsystem. Allowing for the restrictions imposed by servicing all the previous instructions, we need to ensure that one of the N robots in the system will carry out the instruction in such a manner that the servicing time will be minimal. In other words, we must set up a timetable for the movements of, and materials handling by, the robot that will realize the new instruction in the shortest possible time. The solution to this problem consists in successively solving the following three problems:

Problem 1. Find the route of the robot for which the time of arrival at node x_{i_0} is minimal.

Problem 2. Find the route of the robot of Problem 2 that minimizes the time of transfer of a resource to node x_{j_0}.

Problem 3. Find the route along which the robot that has carried out the instructions to transfer a resource from x_{i_0} to x_{j_0} can be sent to its depot $x_{k_0} \in X_3$ in the minimal possible time.

1.4.2.3 The Bellman Equation for Collisionless Movements of Robots in a Transport Network

As noted earlier, the method of solving the above-formulated problems of optimal collisionless movements rests on the fact that these problems are linear in a space of functions with values in an Abelian semi-group (see Section 1.5). On the basis of the physical meaning of the problems on movements of robots in the transportation subsystem of a flexible automatic manufacturing system, we define the respective Abelian semi-group $R_{\oplus} = (R, \oplus)$ as follows. Let $R = \{\varphi, \varphi: R_T \to \{0, 1\}\}$ be the set of all the characteristic functions on the half-line $R_T = \{t, t \geqslant T\}$, where $T \in R^1_+$ is a parameter, $T \geqslant T_0$, whose value in each of Problems 1 to 3 will be specified later. The semi-group operation $\oplus$: $R \times R \to R$ is the operation of taking the pointwise maximum for two elements of R, or $\forall f, g \in R \quad (f \oplus g)(t) = \max (f(t), g(t))$, $t \in R_T$, and the neutral element of the semi-group O_R is a function φ_0 identically equal to zero on R_T $(\varphi_0(t) = 0, t \in R_T)$.

Let us introduce the discrete Bellman equation that describes the collisionless movements of N robots in a given transport network $G = (X, \Gamma)$. We define the following mappings on G

$$\hat{b}\colon X \to R_T \quad (\forall x_i \in X)\,(b(x_i) = b_i = \{t, \chi_{b_i}(t) = 1\}),\ \chi_{b_i} \in R, \quad (1.4.2.1)$$

$$\hat{c}\colon \Gamma \to R_T\,(\forall (x_i, x_j) \in \Gamma)\ (c((x_i, x_j)) = c_{ij} = (t, \chi_{c_{ij}}(t) = 1\}),$$
$$\chi_{c_{ij}} \in R \qquad (1.4.2.2)$$

(here $\chi_{b_i}(t)$ and $\chi_{c_{ij}}(t)$ are the characteristic functions of the given sets b_i and c_{ij}),

$$L\colon \dot{\Gamma} \to \operatorname{End} R_{\oplus}$$
$$(\forall \varphi \in R)\,(\forall (x_i, x_j) \in \Gamma)\,((L((x_i, x_j))\,\varphi)(t)$$
$$= \min\ (\max_{T \leqslant \tau}\ (\varphi(\tau) \min_{\tau \leqslant \theta \leqslant t - \tau_{ij}} \chi_{b_i}(\theta), \chi_{c_{ij}}(t - \tau_{ij})). \qquad (1.4.2.3)$$

The pair of mappings $(\hat{b}, \hat{c})$ given by (1.4.2.1) and (1.4.2.2) specifies. as follows from their definitions, a system of restrictions imposed by the servicing of all the previous instructions concerning the possible states of any one of the N robots, starting at moment $t \geqslant T$. The connections characteristic L of the medium $M = (X, \Gamma, L, R_{\oplus})$ is constructed in such a manner that it ensures collisionless movements of a robot along the arc (x_i, x_j), specifically, if $\varphi(t)$ is the characteristic function of the set of moments of arrival of a certain robot at node x_i, then the image of $\varphi(t)$ resulting from endomorphism (1.4.2.3) has a support that constitutes the set of all the moments of arrival of this robot at node x_j along arc (x_i, x_j), with the motion along (x_i, x_j) occurring without collisions with the other robots. The

simplest way to establish this is to use the definitions of sets b_i and c_{ij} and the isomorphism ($\sim$) of the algebra of the characteristic functions on R_T to the Boolean algebra of subsets of R_T. We can easily see that

$$\max_{T\leqslant\tau}(\varphi(\tau)\cdot\min_{\tau\leqslant\theta\leqslant t}\chi_{b_i}(\theta))=\chi_{\alpha_i}(t)\sim\alpha_i,$$

where α_i is the set of all possible moments of the beginning of movements from node x_i. Accordingly, $\min(\chi_{\alpha_i}(t), \chi_{c_{ij}}(t)) = \chi_{\beta_i}(t) \sim \beta_i = \alpha_i \cap c_{ij}$ is the set of all possible moments of the beginning of movements from node x_j along the arc (x_i, x_j). The translation $\chi_{\beta_i}(t) \to \chi_{\beta_i}(t - \tau_{ij})$ in the isomorphism $\sim$ yields the set of times of arrival at node x_j along the arc (x_i, x_j).

The state of medium M at point x will be understood to be represented by the characteristic function $s(x) \in R$. The support of function $s(x)$ represents the set of possible times of arrival at node x of a free robot. Then in each of the above-mentioned three problems, function s satisfies a Bellman equation with an appropriate right-hand side, or

$$s=Hs\oplus\mathcal{F}_i=1,\ 2,\ 3, \tag{1.4.2.4}$$

where H, the endomorphism in the space of functions $\Lambda = \{s, s: X \to R\}$, is determined by the following formula (see Section 1.5.1):

$$H=[h_{ij}]_{n\times n},\quad n=|X|,$$

$$h_{ij}=\begin{cases} L((x_i,\ x_j)) & \text{if } (x_i,\ x_j)\in\Gamma,\\ \mathbb{O} \text{ if } (x_i,\ x_j)\notin\Gamma. \end{cases}$$

Here $L((x_i, x_j))$ is defined in (1.4.2.3) and $\mathbb{O}$ is the absorbing endomorphism of semi-group $R_\oplus$, $\mathbb{O}(a) = O_R\ \forall\ a \in R$. For each of the above-mentioned three problems, the function $\mathcal{F}_i$, $i = 1, 2, 3$, is defined as follows:

$$\mathcal{F}_1(x)=\begin{cases} \varphi_0,\ x\in X_1\wedge k(x)=0\vee x\in X_2,\\ \chi_{t_k},\ x\in X_1\wedge k(x)\neq 0,\\ \chi_{t(x)},\ x\in X_3\wedge n(x)=0,\\ \chi_T,\ x\in X_3\wedge n(x)\neq 0, \end{cases}$$

where $T = T_0$, $k(x)$ is the number of the first robot to finish servicing its last instruction at time $t_k > T$ at node $x \in X_1$, $k(x) = 0$ if there is not a single robot that finishes servicing the instructions by time t_k at node x, $n(x)$ is the number of robots in depot $x \in X_3$ at time T, $t(x) > T$ is the time of arrival at depot $x \in X_3$ of the first robot to arrive at depot x after servicing its individual instruction (provided that all the previous instructions have been serviced), and χ_a is the characteristic function with support $[a, +\infty]$.

Next,

$$\mathscr{F}_2(x) = \begin{cases} \varphi_0 & \text{if } x \neq x_{i_0}, \\ \chi_T & \text{if } x = x_{i_0}, \end{cases}$$

where $T = T_1$ is the moment of termination of materials handling at node x_{i_0} by the first robot to arrive at node x_{i_0} along the extremal of Problem 1.

Finally,

$$\mathscr{F}_3(x) = \begin{cases} \varphi_0 & \text{if } x \neq x_{j_0}, \\ \chi_T & \text{if } x = x_{j_0}, \end{cases}$$

where $T = T_2$ is the moment of termination of materials handling at node x_{j_0}.

The above formulas for the right-hand sides $\mathscr{F}_i$ of the Bellman equation follow directly from the model of functioning of a transportation subsystem and the definition of the state s of the medium.

The solution to the steady-state Bellman equation (1.4.2.4) in each of the three optimization problems to which the problem of optimal control of robots was earlier reduced is given by the algorithmic procedures to be discussed in Section 1.5. Here we will only give the procedure for reconstructing the optimal trajectories of robot movements.

Let $s = s_i$ be a solution to Eq. (1.4.2.4) with right-hand side $\mathscr{F}_i$ and let x_q be the final node of the extremal μ_i, $i = 1, 2, 3$; $x_q = x_{i_0}$ for μ_1, $x_q = x_{j_0}$ for μ_2, and $x_q = x_k$ for μ_3.

(1) Put $j = q$ and calculate

$$t(x_j) = \min\{t \mid \max_{T \leqslant \tau} (s(x_j)(\tau) \min_{\tau \leqslant \theta \leqslant t} \chi_{b_j}(\theta)) = 1\}.$$

(2) Find x_i from the equation

$$L((x_i, x_j))(s(x_i))(t)|_{t=t(x_j)} = 1. \tag{1.4.2.5}$$

(3) Put

$$\bar{t}(x_i) = t(x_j) - \tau_{ij}. \tag{1.4.2.6}$$

(4) Calculate

$$t(x_i) = \max\{t \mid \max(\chi_{[T, \bar{t}(x_i)]}(t), \max_{T \leqslant \tau} (s(x_i)(\tau) \min_{\tau \leqslant \theta \leqslant t} \chi_{b_i}(\theta))) = 1\}, \tag{1.4.2.7}$$

where

$$\chi_{[T, \bar{t}(x_i)]}(t) = \begin{cases} 1 & \text{if } t \in [T, \bar{t}(x_i)], \\ 0 & \text{if } t \notin [T, \bar{t}(x_i)]. \end{cases}$$

(5) Put $j = i$ and repeat the calculations by formulas (1.4.2.5)-(1.4.2.7) until

$$\forall x_i \in X, \ L((x_i, x_j))(s(x_i))(t)|_{t=t(x_j)} \neq 1.$$

The meaningfulness of this procedure stems from the existence of solution s_i, formulas (1.4.2.5)-(1.4.2.7), and the model of functioning of a system of robots due to which (1) $t(x_q)$ is the minimal possible time of arrival of a robot at final node x_q, (2) solution x_i to Eq. (1.4.2.5) is the node from which this robot traveled to x_j and arrived at x_j at time $t(x_j)$, (3) $\bar{t}(x_i)$ is the time of departure of the robot from node x_i to node x_j, and (4) $t(x_j)$ is the time of arrival of the robot at node x_j. The recursion in item (5) of the algorithm provides in sequence the arcs of extremal μ_i for the solution $s = s_i$. At the end of the algorithm, x_j is the initial node of μ_i from which the robot departs at time $\bar{t}(x_j)$

Remark 1.4.2.1 Note that the solutions of the above subproblems make it possible to rapidly introduce changes in the program controlling the movements of the robots by sending them, after they have finished materials handling, either to the depots or to service new instructions. For practical application of the above process to concrete transportation subsystems one must know the temporal parameters of the functioning of the various parts of such a subsystem. These may be found, for example, by modeling the time of operation of the robots on the basis of the so-called RTM-method (robot, time, movement) [1.66].

Remark 1.4.2.2 The "fuzziness" of the parameters of the problem may be taken into account in the above-discussed scheme of the solution to the problem by considering the semi-group $R_{\oplus} = (R, \oplus)$, where $R = \{\varphi \mid \varphi: R_T \to [0, 1]\}$; here the operation $\oplus$ is defined in a manner similar to the one considered above.

The results obtained here in solving the problem of optimal operational control of a system of robots make it possible to carry out an optimal design of a transport network on the basis of the worked-out procedures of control of transportation robots and the methods of imitation modeling.

1.5 Algorithms for Solving the Generalized Bellman Equation

When the discrete medium is irregular and, hence, no passage to the limit is possible in the corresponding discrete Bellman equation as $n \to \infty$, with $n = |X|$ the number of points of the medium, it is natural to employ numerical methods.

Below (in Section 1.5.2) we introduce sufficiently effective (with polynomial estimates of complexity) algorithms for solving the generalized Bellman equation. The idea of constructing these algorithms stems from the Huygens-Fresnel principle well-known in mathematical physics. This principle provides an essentially algorithmic method for constructing a wavefront of perturbations that

are transmitted in the medium locally [1.67]. In this section we consider the Bellman equation for a restricted discrete medium in a situation that is more general than that discussed in Sections 1.1 and 1.2, namely, we will assume that the connections characteristic L of the medium is an endomorphism of an abstract Abelian semi-group $R_\oplus = (R, \oplus)$, where $\oplus$ is the semi-group operation. If on $(R, \oplus)$ we have defined an additional commutative semi-group operation $\odot$, then, defining endomorphisms as right translations onto R that realize a representation of semi-group $(R, \odot)$, we arrive in this particular case at the Bellman equation in spaces with values in semi-rings considered in the previous sections.

1.5.1 The Generalized Bellman Equation in Spaces with Values in an Abelian Semi-group

Here we introduce the inhomogeneous steady-state Bellman equation in spaces of functions with values in an Abelian semi-group. We find that an analog of Fredholm's first theorem holds true for this equation [1.68].

Let $R_\oplus = (R, \oplus)$ be an abstract Abelian semi-group with the commutative semi-group operation of "addition" $\oplus$: $R^2 \to R$ and the neutral element O_R. It is assumed that on R a partial ordering relation $\leqslant$ and a metric ρ: $R^2 \to [0, \infty]$ are specified that are concordant with each other and the operation $\oplus$ via the axioms considered in Section 1.1.1.

We denote the set of endomorphisms of the Abelian semi-group $R_\oplus$ by $E = \operatorname{End} R_\oplus$, with $E = \{h: R \to R \mid h(a \oplus b) = h(a) \oplus h(b)\}$. Let us consider the semi-ring $\mathscr{E} = (E, \oplus, \odot)$ of endomorphisms equipped with the semi-group operations of "addition" $\oplus$ and "multiplication" (or "product") $\odot$: $\forall\, h, g \in E$, $\forall\, a \in R$, $(h \oplus g)(a) = h(a) \oplus g(a)$, $(h \odot g)(a) = g(h(a))$ and the neutral elements $\mathbb{O} \in E$ and $\mathbb{1} \in E$, respectively: $\mathbb{O}$ is the absorbing endomorphism $\forall\, a \in R$, $\mathbb{O}(a) = O_R$, and $\mathbb{1}$ is the identity endomorphism $\forall\, a \in R$, $\mathbb{1}(a) = a$. It can be easily verified that $\oplus$ is commutative on E: $\forall\, h, g \in E$, $h \oplus g = g \oplus h$ and is related to $\odot$ through the distributivity relations $\forall f, g, h \in E$,

$$f \odot (g \oplus h) = (f \odot g) \oplus (f \odot h),$$

$$(g \oplus h) \odot f = (g \odot f) \oplus (h \odot f).$$

Let $M = (X, \Gamma, L, R_\oplus)$ be a discrete restricted $(|X| = n$ medium with a connections characteristic L: $\Gamma \to \operatorname{End} R_\oplus$ $(|\Gamma| = m)$ that satisfies the condition

$$\forall u \in \Gamma, \quad L(u)(O_R) = O_R. \tag{1.5.1.1}$$

By Λ we denote the space of states of medium M, or $\Lambda = \{\varphi \mid \varphi: X \to R\}$. In Λ we define the structure of the Abelian semi-group equipped

with a commutative operation of "addition" $\oplus$: $\Lambda^2 \to \Lambda$, $\forall \varphi, \psi \in \Lambda$, $\forall\, x \in X$, $(\varphi \oplus \psi)(x) = \varphi(x) \oplus \psi(x)$ and a neutral element $O_\Lambda \in \Lambda$, $O_\Lambda(x) = O_R$. The connections characteristic L of medium M will be extended on $X^2 \setminus \Gamma$ via the absorbing endomorphism $\mathbb{O}$. We denote the resultant mapping $X^2 \to E$ by H:

$$\forall u = (x_i,\ x_j) \in X^2,\quad H(u) = H((x_i,\ x_j)) = h_{ij} = \begin{cases} L(x_i,\ x_j) \text{ if } (x_i,\ x_j) \in \Gamma, \\ \mathbb{O} \text{ if } (x_i,\ x_j) \notin \Gamma. \end{cases} \tag{1.5.1.2}$$

We define the endomorphism H: $\Lambda \to \Lambda$ that acts in the space Λ of states of medium M via the formula

$$\forall \lambda \in \Lambda,\quad (H\lambda)(x_j) = \bigoplus_{i=1}^{n} h_{ij}(\lambda(x_i)) = \bigoplus_{x_i \in \Gamma(x_j)} L(x_i,\ x_j)(\lambda(x_i)),\ x_j \in X, \tag{1.5.1.3}$$

where $\Gamma(x_j) = \{x_i \in X,\ (x_i,\ x_j) \in \Gamma\}$ is a neighborhood of point x_j In other words, endomorphism H (1.5.1.3) is represented by matrix $[h_{ij}]_{n \times n}$ that we will call the operator matrix of the medium.

With endomorphism H (1.5.1.3) we associate the steady-state Bellman equation in the space of functions with values in the abstract Abelian semi-group for the discrete medium M:

$$s = Hs \oplus \mathcal{F}, \tag{1.5.1.4}$$

with $\mathcal{F} \in \Lambda$ a given function $X \to R$.

Our goal is to find a solution to Eq. (1.5.1.4). We assume that the endomorphisms h_{ij} (1.5.1.2) are continuous with respect to uniform convergence in R: $a_n \to b \Leftrightarrow \rho(a_n,\ b) \to 0$, and the function $\mathcal{F}(x)$ is bounded: $\forall\ x \in X$, $O_R \leqslant \mathcal{F}(x) \leqslant \text{const}$. For the generalized Bellman equation (1.5.1.4) we consider the following problem: find all the bounded solutions $s = s(x)$, $\forall\ x \in X$, $O_R \leqslant s(x) \leqslant \text{const}$. We denote the set of all such solutions by B. It has been established that with certain conditions imposed on operator H the set B is not empty and, more than that, an analog of Fredholm's first theorem [1.68] holds true, in other words, the general solution to Eq. (1.5.1.4) can be represented in the form of a sum of an arbitrary solution to the homogeneous equation corresponding to Eq. (1.5.1.4) and a particular solution to the inhomogeneous equation. This particular solution of the initial steady-state problem is defined as the solution to the stabilization Cauchy problem in the space of functions with values in the Abelian semi-group.

The stabilization Cauchy problem corresponding to the steady-state generalized Bellman equation in the space of functions with values in the Abelian semi-group $(R,\ \oplus)$ and associated with the

discrete medium M is formulated as follows: find

$$\lim_{t\to\infty} u(x, t) = u^*(u_0), \tag{1.5.1.5}$$

where $u(x, t)$ is a bounded solution to the Cauchy problem for the inhomogeneous equation

$$u_{t+1} = Hu_t \oplus \mathcal{F} \tag{1.5.1.6}$$

with the initial condition

$$u\,|_{t=0} = u_0, \quad \forall\, x \in X, \; O_R \leqslant u_0(x) \leqslant \text{const}. \tag{1.5.1.7}$$

The limit $u^*(u_0)$ is called the solution to the stabilization Cauchy problem. By $\hat{R}_t(u_0)$ we denote the resolving operator of problem (1.5.1.5)-(1.5.1.7), and by $\hat{B}_{u_0}$ the resolving operator of the stabilization Cauchy problem: $\hat{B}_{u_0} = \lim\limits_{t\to\infty} \hat{R}_t(u_0)$, with u_0 fixed. Here and in what follows the limit is understood in the sense of strong convergence of operators on the subspace of functions that are bounded and belong to Λ.

The structure of the semi-ring of endomorphisms End $R_\oplus$ of the Abelian semi-group $R_\oplus = (R, \oplus)$ induces in a natural way the structure of the semi-ring on the set of endomorphisms of semi-group $\Lambda_\oplus = (\Lambda, \oplus)$. The respective semi-group operations of "addition" $\oplus$ and "multiplication" $\odot$ are defined by the formulas

$$\forall\, A, \; B \in \text{End}\, \Lambda_\oplus, \quad \forall\, \lambda \in \Lambda,$$

$$(A \oplus B)(\lambda) = (A\lambda) \oplus (B\lambda), \; (A \odot B)(\lambda) = B(A(\lambda)),$$

which in the matrix representation of endomorphisms have the form

$$A = [a_{ij}]_{n\times n}, \quad B = [b_{ij}]_{n\times n},$$

$$A \oplus B = [a_{ij} \oplus b_{ij}]_{n\times n},$$

$$A \odot B = \left[\bigoplus_{k=1}^{n} a_{ik} \odot b_{kj}\right]_{n\times n}.$$

By H^t we denote the operator H raised to power t, that is, $H^0 = D = [\delta_{ij}]_{n\times n}$, $H^t = H^{t-1} \odot H$, $t \geqslant 1$, where $\delta_{ij} = \delta(i - j)$ is the generalized delta function in the semi-ring $\mathscr{E} = (E, \oplus, \odot)$: $\delta_{ij} = \mathbb{1}$ if $i = j$ and $\delta_{ij} = \mathbb{O}$ if $i \neq j$ [1.69]. Then $H^{(t)} = \bigoplus\limits_{k=0}^{t} H^k$.

Assertion 1.5.1.1 *Suppose that the limits* $\lim\limits_{t\to\infty} H^t = H^\infty$ *and* $\lim\limits_{t\to\infty} H^{(t)}$ H^* *exist. Then the operator* B_{u_0}: $\Lambda \to \Lambda$ *exists and can be represented in the form*

$$\hat{B}_{u_0} - H^\infty u_0 \oplus H^*. \tag{1.5.1.8}$$

Proof. For a finite t, in view of the linearity of H, the solution to problem (1.5.1.5)-(1.5.1.7) can be represented in the form $u(x, t) = H^t u_0 \oplus f_{t-1}$, where f_t is the solution to Eq. (1.5.1.6) with a zero initial condition. By Duhamel's theorem,

$$f_t = \overset{\oplus}{\int\limits_{[0,\, t]}} H^{t-\tau}\mathcal{F}\, d\tau$$

$$= \bigoplus_{0 \leqslant \tau \leqslant t} H^{t-\tau}\mathcal{F} = \bigoplus_{k=0}^{t} H^k \mathcal{F} = H^{(t)}\mathcal{F}.$$

Combining this with the previous equality and passing to the limit as $t \to \infty$ we get (1.5.1.8).

Remark 1.5.1.1 Generally the operators $\hat{R}_t(u_0)$ and $\hat{B}_{u_0}$ are not endomorphisms acting in Λ if operation $\oplus$ is not idempotent.

Corollary *The fact that endomorphism H is continuous and* (1.5.1.8) *imply that for each u_0 the function $s = \hat{B}_{u_0}\mathcal{F} = H^\infty u_0 \oplus H^*\mathcal{F}$ is a solution to the steady-state Bellman equation* (1.5.1.4), *with $H^\infty u_0 = g_0$ a solution to the homogeneous equation $Hg = g$. We denote the solution to the inhomogeneous equation (i.e. with $H^*\mathcal{F}$) by s^* and call it the Duhamel solution. Obviously, $s^* = \hat{B}_{0_\Lambda}\mathcal{F}$. Thus, with the above-formulated restrictions imposed on H, the steady-state equation* (1.5.1.4) *has among its solutions the solution to the stabilization Cauchy problem, $s = \hat{B}_{u_0}\mathcal{F} = H^\infty u_0 \oplus s^*$.*

Theorem 1.5.1.1 *Suppose the conditions of the assertion* 1.5.1.1 *are met. Then on the set B of all bounded solutions to the steady-state Bellman equation, or $\forall\, s \in B$, we have*

$$s = g \oplus s^*, \tag{1.5.1.9}$$

where g is a bounded solution to the homogeneous equation $Hg = g$, and $s^* = \hat{B}_{u_0}\mathcal{F}$ is the Duhamel solution of the stabilization Cauchy problem (1.5.1.5)-(1.5.1.7).

Proof. Consider a chain of equalities following from Eq. (1.5.1.4) and the definition of $H^{(t)}$. For every finite $k \in \mathbb{Z}_+$ we have

$$\begin{aligned} s &= Hs \oplus \mathcal{F} \Rightarrow s = H(Hs \oplus \mathcal{F}) \oplus \mathcal{F} \\ &= H^2 s \oplus H\mathcal{F} \oplus D\mathcal{F} = H^2 s \oplus H^{(1)}\mathcal{F} \Rightarrow \ldots \Rightarrow s \\ &= H^k s \oplus H^{(k-1)}\mathcal{F}. \end{aligned}$$

Passing to the limit, as $k \to \infty$, in the last equality and putting $g = H^\infty s$, we get (1.5.1.9).

For optimization problems in which operation $\oplus$ is idempotent, the solutions to the generalized Bellman equation (1.5.1.4) possess the following properties:

Lemma 1.5.1.1 *Let the semi-group operation* $\oplus$ *be idempotent,* $\forall\, a \in R,\ a \oplus a = a$, *and let the strong limits* $\lim\limits_{t\to\infty} H^t = H^\infty$ *and* $\lim\limits_{t\to\infty} H^{(t)} = H^*$ *exist. Then the following assertions are true*:

(1) $\forall u_0 \in \Lambda\ \hat{B}_{u_0}$ *is an endomorphism in* $\Lambda_\oplus$, *with* $\hat{B}_{u_0} \odot \hat{B}_{u_0} = \hat{B}_{u_0}$;

(2) *the Duhamel solution* $s^* = \hat{B}_{O_\Lambda}\mathcal{F}$ *is the neutral element on the set* B *of all bounded solutions to Eq.* (1.5.1.4), *or* $\forall\ s \in B\quad s \oplus s^* = s$;

(3) $\hat{B}_f = \hat{B}_{O_\Lambda}$ *for every* f *satisfying the condition* $f \oplus s^* = s^*$; *and*

(4) *if* $\tilde{H} = D \oplus H$, *then the Duhamel solutions to the equations* $s = Hs \oplus \mathcal{F}$ *and* $s = \tilde{H}s \oplus \mathcal{F}$ *coincide.*

Proof. The proof follows directly from the previous formulas and the fact that $\oplus$ is idempotent.

1.5.2 The H-method of Numerical Solution of the Generalized Bellman Equation

In this section we discuss two schemes for the numerical solution of the Bellman equation in a space with values in an Abelian semi-group: the Picard method of successive approximations, and (in the case of idempotency of the semi-group operation $\oplus$) the more effective H-method of successive approximations.

Let us consider exact numerical algorithms for constructing the Duhamel solution to the generalized Bellman equation.

1.5.2.1 Picard Method of Successive Approximations

Let $s^P = \{s_t,\ t = 0, 1, 2, \ldots\}$ be the sequence of approximations fixed by the following recursion scheme:

$$s_0 = \mathcal{F},\ s_{t+1} = Hs_t \oplus \mathcal{F},\ t = 0, 1, 2, \ldots . \qquad (1.5.2.1)$$

The sequence stabilizes if $\exists\ t_0^P \in \mathbb{Z}_+$ and $\forall\ t > t_0^P,\ s_{t+1} = s_t$. It is obvious that sequence s^P given by (1.5.2.1) stabilizes if and only if

$$\exists\ t_0^P \in \mathbb{Z}_+,\ \forall\ t \geqslant t_0^P,\ H^{(t)} = H \odot H^{(t)} \oplus D. \qquad (1.5.2.2)$$

In this case $s_{t_0}^P$ is the Duhamel solution s^* to Eq. (1.5.1.4), with the complexity estimate T of constructing s_{t_0} being $n^2 t_0^P\,(T_h + T_\oplus)$, where T_h is the complexity estimate of calculating the endomorphism h_{ij}, and $T_\oplus$ is the complexity estimate of performing operation $\oplus$.

Below we give some conditions on the medium M and the endomorphisms h_{ij} that are sufficient for stabilization of the successive approximations in the Picard method:

(1) $Z(M) = \varnothing$, where $Z(M)$ is the set of all elementary contours of graph $G = (X, \Gamma)$;

(2) $\forall\, h \in \operatorname{Im} L,\ \mathbb{1} \oplus h = \mathbb{1}$;

(3) $\forall\, \nu \in Z(M),\ \mathbb{1} \oplus L(\nu) = \mathbb{1}$; where $\nu = \{u_1, u_2, \ldots, u_k\}$ and $L(\nu) = L(u_1) \odot L(u_2) \odot \ldots \odot L(u_k)$ is a composition of endomorphisms taken along contour ν;

(4) $\exists\, k_0 \in \mathbb{Z}_+$ such that for every family (i.e. a collection of elements not necessarily pairwise distinct) of endomorphisms $h_1, h_2, \ldots, h_{k_0}$ belonging to $\operatorname{Im} L$ we have $\mathbb{1} \oplus (h_1 \odot h_2 \odot \ldots \odot h_{k_0}) = \mathbb{1}$.

When the endomorphisms in $\operatorname{Im} L$ are commutative, the conditions (2)-(4) may be made weaker:

(5) $\exists k_0 \in \mathbb{Z}_+,\quad \forall h \in \operatorname{Im} L,\quad h^{(k_0)} = h^{(k_0+1)} = h \odot h^{(k_0)} \oplus \mathbb{1}$, where $h^{(t)} = \bigoplus_{k=0}^{t} h^t,\ t \in \mathbb{Z}_+,\ h^0 = \mathbb{1}$;

(6) $\exists k_0 \in \mathbb{Z}_+,\ \forall \nu \in Z(M),\ L^{(k_0)}(\nu) = L^{(k_0+1)}(\nu)$.

Let us prove, say, the sufficiency of condition (6). (The proof for conditions (1)-(5) is similar.) We introduce $\lambda_t = H^t \mathcal{F}$, $t = 0, 1, 2, \ldots$, and represent λ_t in the form of a discrete Feynman "continual integral", an integral along paths [1.22, 1.23]. For each $t = 0, 1, 2, \ldots$ we define a set of points of the medium, $\Phi_t \subseteq X$, through the relationship $\Phi_t = \{x \in X \mid \lambda_t(x) \neq O_R\}$. Then, in view of condition (1.5.1.1) and formula (1.5.1.2), we have

$$\lambda_i(y) = (H\lambda_{t-1})(y) = \bigoplus_{x \in \Phi_{t-1}} \ \bigoplus_{(x,\, y) \in \Gamma} L((x,\, y))\,(\lambda_{t-1}(x)). \tag{1.5.2.3}$$

Enforcing recursion in (1.5.2.3) to $t = 0$, we get

$$\lambda_t(y) = (H^t.\mathcal{F})(y) = \bigoplus_{x \in \Phi_0} \ \bigoplus_{\mu \in M^t_{x \to y}} L(\mu)\,(\mathcal{F}(x)). \tag{1.5.2.4}$$

Here $\Phi_0 = \{x \in X \mid \lambda_0(x) = H^0.\mathcal{F} = .\mathcal{F} \neq O_R\}$, $M^t_{x\to y}$ is the set of routes $\mu = \{u_1, u_2, \ldots, u_t\}$ of length $t = |\mu|$ connecting points x and y of the medium, $L(\mu) = \bigodot_{i=0}^{t} L(u_i) = L(u_1) \odot L(u_2) \odot \ldots \odot L(u_t)$ is the composition of endomorphisms along route μ (the characteristic of route μ). Formula (1.5.2.4) constitutes a representation of $H^t.\mathcal{F} = \lambda_t$ in the form of a discrete Feynman "continual integral," the integral along routes μ in M. By direct substitution into (1.5.2.1) we can easily verify that each member s_t of sequence s^Γ can be represented in the form $s_t = \bigoplus_{k=0}^{t} \lambda_k$. Hence, by virtue of (1.5.2.4), we finally arrive at a representation of s_t in the form of a "sum" along routes:

$$s_t(y) = (H^{(t)}.\mathcal{F})(y) = \bigoplus_{x \in \Phi_0} \ \bigoplus_{\mu \in M^{(t)}_{x \to y}} L(\mu)\,(\mathcal{F}(x)). \tag{1.5.2.5}$$

Here $M_{x\to y}^{(t)} = \bigcup_{k=0}^{t} M_{x\to y}^{t}$ is the set of all routes $\mu = \{u_1, u_2, \ldots, u_k\}$ from x to y whose lengths are no greater than t ($|\mu| \leqslant t$), and $L(\mu) = \mathbb{1}$ at $|\mu| = 0$.

Any route μ, $|\mu| = k \in \mathbb{Z}_+$, can be represented in the form $\mu = \mu_0 \cup \bigcup_{p=1}^{r(\mu_0)} \nu_p^{\alpha_p}$, where $\mu_0 \in \overline{M}_{x\to y}$ is the elementary route from x to y, $\overline{M}_{x\to y}$ is the set of all elementary routes from x to y, $\nu_p \in Z(M)$ is an elementary contour incident to μ_0, $\alpha_p \in \mathbb{Z}_+$ is the multiplicity of ν_p in μ, and $r(\mu_0)$ is the number of all the elementary coutours incident to μ_0, $\forall \mu_0 \in \overline{M}_{x\to y}$, $r(\mu_0) \leqslant |Z(M)|$, with $k = |\mu| = |\mu_0| + \sum_{p=1}^{r(\mu_0)} |\nu_p| \alpha_p$. With μ_0 we associate the set $M_t(\mu_0) = \{\mu,\ \mu = \mu_0 \cup \bigcup_{p=1}^{r(\mu_0)} \nu_p^{\alpha_p},\ k = |\mu| \leqslant t\}$.

By virtue of the commutativity of $\odot$ and the distributivity of $\odot$ with respect to $\oplus$, we have

$$L_t(\mu_0) = \bigoplus_{\mu \in M_t(\mu_0)} L(\mu) = \Big(\bigodot_{p=1}^{r(\mu_0)} L^{(t)}(\nu_p)\Big) \odot L(\mu_0), \quad (1.5.2.6)$$

where $L^{(t)}(\nu_p) = \bigoplus_{k=0}^{t} L^k(\nu_p)$.

Let $k^* = (n-1) + |Z(M)|\, nk_0$. By virtue of condition (6) and Eq. (1.5.2.6), the latter equality yields $L_{t+1}(\mu_0) = L_t(\mu_0)$ for $t \geqslant k^*$. Combining this with (1.5.2.5), we obtain the following chain of equalities:

$$\begin{aligned} s_t(y) &= \bigoplus_{x\in\Phi_0} \bigoplus_{\mu \in M_{x\to y}^{(t)}} L(\mu)(.\mathcal{F}(x)) \\ &= \bigoplus_{x\in\Phi_0} \bigoplus_{\mu_0 \in \overline{M}_{x\to y}} \bigoplus_{\mu \in M_t(\mu_0)} L(\mu)(.\mathcal{F}(x)) \\ &= \bigoplus_{x\in\Phi_0} \bigoplus_{\mu_0 \in \overline{M}_{x\to y}} L_t(\mu_0)(.\mathcal{F}(x)). \end{aligned}$$

This, obviously, implies that $s_{t+1} = s_t$ for $t \geqslant k^*$ and, hence, the quantity t_0^{P} required by the stabilization condition (1.5.2.2) does indeed exist, with $t_0^{\mathrm{P}} \leqslant k^*$.

If the operation $\oplus$ is idempotent, $\forall\, a \in R$, $a \oplus a = a$, we have a method of successive approximations for solving the generalized Bellman equation (1.5.1.4) that is more effective than the Picard method.

1.5.2.2 The H-method of Successive Approximations

We denote by π_{ij}: $\Lambda \to \Lambda$ an endomorphism belonging to End $\Lambda_\oplus$ that is defined by the formula

$$\forall s \in \Lambda, \ (\pi_{ij}s)(x) = \begin{cases} s(x) & \text{if} \quad x \neq x_j \\ s(x_j) \oplus h_{ij}(s(x_i)) & \text{if} \quad x = x_j, \end{cases} \tag{1.5.2.7}$$

with h_{ij} defined in (1.5.1.2).

Suppose that $s^H = \{s_0, s_1, s_2, \ldots, s_t, \ldots\}$ is a sequence of approximations fixed by the following recursion scheme:

$$s_0 = \mathcal{F}, \ s_t = \begin{cases} \pi_{ij}s_{t-1} & \text{if} \quad Q_t \neq \varnothing \wedge (i, j) \in Q_i, \\ s_{t-1} & \text{if} \quad Q_t = \varnothing, \end{cases} \tag{1.5.2.8}$$

with

$$Q_t = \{(i, j), \ s_{t-1} \neq \pi_{ij}s_{t-1}\}. \tag{1.5.2.9}$$

Theorem 1.5.2.1. *If any one of the above conditions* (1)-(6) *is met, the sequence* s^H *stabilizes*: $\exists t_0^H \in \mathbb{Z}_+$, $S_{t_0^H} = s_{t_0^H+1}$, with $s_{t_0^H} = s^*$, *where* s^* *is the Duhamel solution to Eq.* (1.5.1.4).

Proof. The proof can be carried out in the same way as in Section 1.5.2.1 and is based on the "continual" representation of s_t (see [1.22, 1.24]).

By virtue of formulas (1.5.2.8) and (1.5.2.9), the sequence $s^H = \{s_t, t = 0, 1, 2, \ldots\}$ that converges, as $t \to t_0^H$, to solution s^* of Eq. (1.5.1.4) is not defined in a unique manner. Depending on which of the sufficiency conditions (1)-(6) is met, there are different methods (H-algorithms) of constructing the sequence $s_t^H \to s^*$, with complexity estimates that are better than in the Picard method. The reason is that for the idempotent operation $\oplus$ the order in which the endomorphism π_{ij} (1.5.2.7) is chosen from the set of all admissible endomorphisms defined by the sets Q_t (1.5.2.9) is not important. The choice of a definite order in which the points and connections in medium M are selected in an H-method of successive approximations, in other words, the choice based on an analysis of the topology and properties of the local transformations h_{ij}, yields concrete H-algorithms with different complexity estimates.

Algorithm A_H^1:

Put $s \neq \mathcal{F}$, $Q = \{x \in X, \mathcal{F}(x) \neq O_R\}$;
while $Q \neq \varnothing$ **do**
  select point $x_i \in Q$;
  $Q = Q \setminus \{x_i\}$;
  for $x_j \in \Gamma(x_i)$ **do**
    $R = \pi_{ij}s$;
    if $s \neq R$ **then** $s = R$; $Q = Q \cup \{x_i\}$ **fi**
  od
od.

This algorithm can be applied when any one of the sufficiency conditions (1)-(6) is met. Let us estimate the accuracy of algorithm A^1_H when, say, condition (2) is satisfied, assuming that the semi-group operation $\oplus$ is induced by the linear ordering relation ρ (see Section 1.5.3 below). Refining the choice of point $x_i \in Q$ in algorithm A^1_H in such a manner that the point satisfies the condition $s(x_i) = \bigoplus_{x \in Q} s(x)$ and representing medium M by the continuity structure [1.70] for algorithm A^1_H, we find that the complexity estimate is $n^2 T_\oplus + mT_h$. Under the same conditions but with a representation of set Q in the form of a sorting tree with respect to ρ, the complexity estimate for A^1_H is equal to $m \log nT_\oplus + mT_h$ [1.24]. If medium M satisfies the sufficiency condition (1), there is an H-algorithm that is more effective than A^1_H, namely,

Algorithm A^2_H:
for $x \in X$ **do**
  $sT(x) = |\{y \in X,\ (y, x) \in \Gamma\}|$;
od
  put $s = \mathcal{F}$; $Q = \{x \mid sT(x) = 0\}$;
while $Q \neq \varnothing$ **do**
  select point $x_i \in Q$; $Q = Q \setminus \{x_i\}$;
  for $x_j \in \Gamma(x_i)$ **do**
    $s = \pi_{ij} s$;
    $sT(x_j) = sT(x_j) - 1$;
    if $sT(x_j) = 0$ **then** $Q = Q \cup \{x_j\}$ **fi**
  od
od.

The complexity estimate for A^2_H is equal to $m(T_h + T_\oplus)$.

1.5.3 Examples of Solution of Extremal Problems of Discrete Optimization Linear in Spaces with Values in an Abelian Semi-group

In this section we consider the solutions of single-iteration and multi-iteration problems as examples of solution of extremal problems.

The concept of a generalized Bellman equation in spaces with values in an Abelian semi-group makes it possible to give the statement of, and the solution methods for, a broad class of problems of discrete optimization on graphs. Here we restrict our discussion to the special Abelian semi-group $R_\oplus = (R, \oplus)$ in which the semi-group operation $\oplus$ is induced by the linear ordering relation ρ

on R:

$$a \oplus b = \begin{cases} a \text{ if } a\rho b, \\ b \text{ if } \neg\, a\rho b, \end{cases} \qquad (1.5.3.1)$$

The choice of such a semi-group is determined, for one, by the fact that all known classical problems of optimization on graphs [1.37, 1.70, 1.71], with appropriate choice of the discrete medium, prove to be linear in spaces of functions with values in this semi-group and, hence, a numerical solution for these problems can be obtained by employing the H-algorithms discussed in Section 1.5.2.

From the standpoint of the generalized Bellman equations in the above-noted semi-group, problems of discrete optimization on graphs divide naturally into two classes: (1) problems whose solution can be found by solving the Bellman equation in a fixed medium (single-iteration problems), and (2) problems whose solution requires constructing consecutively solutions of the Bellman equation in a medium that changes from iteration to iteration (multi-iteration problems).

1.5.3.1 Single-iteration Problems

As an example of single-iteration problems we consider the problems of so-called discrete optimization on graphs of the trajectory type. The statement of these problems is naturally related to the continual representation of the solution to the generalized Bellman equation. Specifically let s^* be the (Duhamel) solution to Eq. (1.5.1.4), $s^* = \lim\limits_{t\to\infty} s_t$. Then, in the limit of $t \to \infty$ Eq. (1.5.2.5) yields the following "continual" representation for solution s^*:

$$\forall y \in X,\ s^*(y) = \bigoplus_{x\in\Phi_0} \bigoplus_{\mu\in M_{x\to y}} L(\mu)(\mathscr{F}(x)), \qquad (1.5.3.2)$$

where $M_{x\to y} = \lim\limits_{t\to\infty} M^{(t)}_{x\to y}$ is the set of all routes connecting points x and y of the medium, and $\Phi_0 = \{x \in X,\ \mathscr{F}(x) \neq O_R\}$. By virtue of the definition (1.5.3.1) of operation $\oplus$ formula (1.5.3.2) implies that for every vertex $y \in X$ there is a route μ_0 from point x_0 to point y, $x_0 \in \Phi_0$ for which the following holds true:

$$L(\mu_0)(\mathscr{F}(x_0)) = s^*(y).$$

In other words, in terms of the linear ordering relation ρ (1.5.3.1), route μ_0 is optimal on the set of all routes $M_{x\to y}$, that is, for this route we have

$$\forall\ x \in \Phi_0,\ \forall\ \mu \in M_{x\to y},\ L(\mu_0)(\mathscr{F}(x_0))\ \rho L(\mu)(\mathscr{F}(x)).$$

This naturally leads to the following statement of the problem of the trajectory type in the discrete medium $M = (X, \Gamma, L, R_\oplus)$:

Let an arbitrary element $\mathcal{F}_0 \in \Lambda$ and an arbitrary subset $\Phi \subseteq X$ be given, let $\Phi_0 = \{x \in X, \mathcal{F}_0(x) = O_R\}$, and let $M_{\Phi_0\to\Phi}$ be the set of all the routes connecting the points of sets Φ_0 and Φ, $M_{\Phi_0\to\Phi} = \bigcup_{x\in\Phi_0}\bigcup_{x\in\Phi} M_{x\to y}$. We wish to establish at least one route $\mu_0 \in M_{\Phi_0\to\Phi}$ from $x_0 \in \Phi_0$ to $y_0 \in \Phi$ that satisfies the following ordernig relation (μ_0 is the optimal route):

$$\forall\mu \in M_{\Phi_0\to\Phi}, \quad L(\mu_0)(\mathcal{F}_0(x_0))\,\rho L(\mu)(\mathcal{F}_0(x)). \qquad (1.5.3.3)$$

Note that in view of the above reasoning, the "length" of the optimal route, $L(\mu_0)(\mathcal{F}_0(x_0))$, and the final point of the route, y_0, can be determined uniquely from the solution s^* to the Bellman equation (1.5.1.4) with the right-hand side $\mathcal{F} = \mathcal{F}_0$ through the following formulas:

$$s^*(y_0) = \bigoplus_{y\in\Phi} s^*(y) = L(\mu_0)(\mathcal{F}_0(x_0)).$$

Generally, describing the procedure of reconstructing the route μ_0 from the known solution s^* to Eq. (1.5.1.4) requires drawing on additional constructions [1.24, 1.56]. Here we describe the procedure in a situation when the discrete medium M and the connections characteristic satisfy the sufficiency conditions (1)-(3), in view of which route μ_0 is a path. We introduce the notation $\mu_0 = \{x_{p_1}, x_{p_2}, \ldots, x_{p_k}\}$. The nodes x_{p_j}, $j = 1, \ldots, k$, can be found through the following recursion scheme

$$\begin{aligned}
&x_{p_k} = y_0,\\
&s^*(x_{p_i}) = h_{p_{i-1}p_i}(s^*(x_{p_{i-1}})),\ i = k,\ k-1, \ldots, 2,\\
&\forall x_q \in X,\ s^*(x_{p_1}) \neq h_{qp_1}(s^*(x_q)),\\
&x_{p_1} = x_0.
\end{aligned}$$

Note that the optimization problems of the trajectory known from the literature [1.70] are "included" in problem (1.5.3.3) under the following choice of the endomorphisms h_{ij}. Specifically, we will assume that on $(R, \oplus)$ we have defined an additional semi-group operation of multiplication $\odot$. Then we define h_{ij} as the right translations on R that realize the representation of semi-group $(R, \odot)$, that is, $\forall a \in R$, $h_{ij}(a) = a \odot f_{ij}$, where $f_{ij} \in R$. In this case Eq. (1.5.1.4) is the Bellman equation in the space of functions with values in the semi-ring $A = (R, \oplus, \odot)$. If we put $\mathcal{F}_0 = \delta_x^\zeta$ in (1.5.3.3), with δ_x^ζ the "delta function" in semi-ring A, or

$$\delta_x^\zeta = \delta^\zeta(x) = \begin{cases} 1_R & \text{if } x = \zeta, \\ O_R & \text{if } x \neq \zeta \end{cases} \quad \text{and } \Phi = \{y\},$$

we arrive at the problem of constructing the optimal path between two vertices, ζ and y, in graph $G = (X, \Gamma)$. Table 1.5.3.4 lists

examples of semi-rings and complexity estimates for H-algorithms for solving problem (1.5.3.3). For similar structures see [1.74].

Table 1.5.3.4

Nos.	R	$a \oplus b$	O_R	$a \odot b$	1_R	Complexity
1	R^1	$\min(a, b)$	$+\infty$	$a+b$	0	$O(n^2)$, $O(m \log n)$
2	R^1_+	$\max(a, b)$	0	$a+b$	0	$O(m)$
3	R^1	$\min(a, b)$	$+\infty$	$\max(a, b)$	$-\infty$	$O(n^2)$, $O(m \log n)$
4	R^1_+	$\max(a, b)$	0	$\min(a, b)$	$+\infty$	$O(n^2)$, $O(m \log n)$
5	$[0, 1]$	$\min(a, b)$	0	$a \times b$	1	$O(n^2)$, $O(m \log n)$
6	$[0, 1]$	$\max(a, b)$	1	$a \times b$	1	$O(n^2)$, $O(m \log n)$
7	$\{0, 1, \ldots, n\}$	$\min(a, b)$	n	$\min(n, a+b)$	0	$O(m)$

As another example of a problem of the trajectory type (1.5.3.3), let us take the problem of constructing the shortest path through a network with a variable transit time along the arcs [1.72]. Suppose that to each arc u of graph $G = (X, \Gamma)$ we assign a nonnegative function $\varphi_u(t)$. The independent variable t is the starting time of motion from node x_i along arc $u = (x_i, x_j)$, while the value of $\varphi_u(t)$ at time t is the transit time along arc u. We wish to construct the shortest (namely, "the fastest") route $\mu_0^{t_0}$ from a specified node $x_0 \in X$ to any other node of the graph, $x \in X$, under the condition that the departure from node x_0 occurs at time t_0. This problem has been solved in [1.73] by the dynamic programming method employing a discrete time scale, with the estimates of the temporal and capacity complexities of calculations and the accuracy of the resulting solution depending essentially on the step on the time scale.

A more effective method of solving this problem with a complexity estimate equal to $O(n^2)$ or $O(m \log n)$ ($n = |X|$ and $m = |\Gamma|$) [1.24, 1.56] is based on reducing it to the solution to the generalized Bellman equation (1.5.1.4). The discrete medium $M = (X, \Gamma, L, R_\oplus)$ corresponding to this problem is defined in the following manner: (1) $G = (X, \Gamma)$ representing the initial network, (2) $R_\oplus = (R, \oplus)$, where $R = \{t \in R^1, t \geqslant t_0\}$ and $\oplus$ = mni, and (3) the endomorphisms $L(u) \in \operatorname{End} R_\oplus$ are specified by the formula $\forall a \in R$, $L(u) \times (a) = \min_{\tau \geqslant a} \{\tau + \varphi_u(\tau)\}$. Then we can easily show that the length $s^*(x)$ of the shortest route $\mu_0^{t_0}$ from $x_0 \in X$ to $x \in X$ is the solution to the Bellman equation (1.5.1.4) with the right-hand side $\mathscr{F}$ of the form

$$\mathscr{F}(x) = \begin{cases} t_0 & \text{if } x = x_0, \\ O_R = \infty & \text{if } x \neq x_0. \end{cases}$$

1.5.3.2 Multi-iteration Problems

The state of, and solution methods for, the problems given below are discussed from the standpoint of the Bellman equation (1.5.1.4) in the space of functions with values in the semi-ring $(R, \oplus, \odot)$ supplemented by an additional condition, the commutativity of the "multiplication" operation $\odot$.

The generalized assignment problem. Suppose that $G = (X, \Gamma)$ is an arbitrary nonoriented graph. By $\mathcal{P} = \{P\}$ we denote the set of all combinations of pairs P of edges in graph G, and by $n(P)$ the number of edges entering P. Then $\mathcal{P}_0 = \{P_0 \in \mathcal{P} \mid \forall P \in \mathcal{P}, n(P_0) \geqslant n(P)\}$ is the set of all combinations of pairs of maximal power. On the set of edges of the graph we specify a weighting function φ with values in R, $\varphi \colon \Gamma \to R$. We define the weight $\varphi(P)$ of the combination of pairs, P, thus: $\varphi(P) = \bigodot_{u \in P} \varphi(u)$. We wish to find the combination of pairs $P_0 \in \mathcal{P}_0$ for which $\varphi(P_0) = \bigoplus_{P \in \mathcal{P}_0} \varphi(P)$.

At each ith iteration, $i = 0, 1, 2, \ldots, n(P_0) - 1$, of the algorithm for constructing the optimal combination of pairs, P_0, the following sequence of steps is carried out:

(1) by the combination of pairs P constructed in the previous iteration we determine the medium $M^i = (X, \Gamma, L^i, R_\oplus)$:

$$\forall a \in R,\ L^i(u)(a) = \varphi^i(u) \odot a,$$
$$\varphi^i(u) = \begin{cases} \varphi(u) & \text{if } u \in \Gamma \setminus P, \\ \bar{\varphi}(u) & \text{if } u \in P, \end{cases} \tag{1.5.3.4}$$

where $\bar{\varphi}(u)$ is listed in Table 1.5.3.5,

Table 1.5.3.5

No.	Assignment	R	$a \oplus b$	0_R	$a \odot b$	1_R	$\bar{\varphi}(u)$
1	Maximal power	$\{-1, 0, +1\}$	$\max(a, b)$	-1	$\begin{cases} a+b, & a \neq b \\ a, & a = b \end{cases}$	0	-1
2	Minimal weight	R^1	$\min(a, b)$	$+\infty$	$a+b$	0	$-\varphi(u)$
3	Maximal weight	R^1	$\max(a, b)$	$-\infty$	$a+b$	0	$-\varphi(u)$
4	Minimax	R^1	$\min(a, b)$	$+\infty$	$\max(a, b)$	$-\infty$	$\varphi(u)$
5	Maximin	R^1	$\max(a, b)$	$-\infty$	$\min(a, b)$	$+\infty$	$\varphi(u)$
6	Least reliable	R^1_+	$\min(a, b)$	$+\infty$	$a \times b$	1	$\varphi^{-1}(u)$
7	Most reliable	R^1_+	$\max(a, b)$	0	$a \times b$	1	$\varphi^{-1}(u)$

(2) in medium M^i we solve the Bellman equation corresponding to the problem of determining the optimal alternating chain μ_0, that is, a chain for which

$$\bigodot_{u \in \mu_0} \varphi^i(u) = \bigoplus_{\mu \in \mathcal{M}} \bigodot_{u \in \mu} \varphi^i(u),$$

where $\mathcal{M}$ is the set of alternating chains μ [1.73], and

(3) as is known, along an alternating chain the power of combination of pairs P increases by unity.

The optimality of the algorithm constructed here for the combination of pairs P ($P = P_0$) follows from the fact that at each ith iteration, $i = 1, 2, \ldots, n(P_0) - 1$, for the constructed running combination of pairs P we have

$$\varphi(P) = \bigoplus_{Q \in \mathcal{P}^i} \varphi(Q),$$

where $\mathcal{P}^i = \{P \in \mathcal{P} \mid n(P) = i\}$ is the set of combinations of pairs of power i.

In Table 1.5.3.5 we list, for a number of known optimal assignment problems in a weighted graph, the corresponding semi-rings and weighting functions φ^i, which according to formula (1.5.3.4) determine the connections characteristic L^i of medium M^i.

The generalized transportation problem. Suppose that on the arcs Γ of graph $G = (X, \Gamma)$ we have specified a function $C: \Gamma \to R^1_+$ called the carrying capacity. The graph contains two isolated subsets of nodes: sources $X_1 \subset X$ and sinks $X_2 \subset X$, with $X_1 \cap X_2 = \varnothing$. The nonnegative function $f: \Gamma \to R^1_+$ that satisfies the relationships $\forall\ u \in \Gamma,\ f(u) \leqslant C(u)$; $\forall\ x \in X_1, \operatorname{div}_f(x) \geqslant 0$; $\forall\ x \in X_2$, $\operatorname{div}_f(x) \leqslant 0$; $\forall\ x \in X \setminus (X_1 \cup X_2)$, $\operatorname{div}_f(x) = 0$, where

$$\operatorname{div}_f(x_i) = \sum_{(x_i, x_j) \in \Gamma} f(x_i, x_j) - \sum_{(x_j, x_i) \in \Gamma} f(x_j, x_i)$$

is said to be the flow [1.44], while $V(f) = \sum_{x \in X_1} \operatorname{div}_f(x) = -\sum_{x \in X_2} \operatorname{div}_f(x)$ is said to be the magnitude of the flow. By $\mathcal{G} = \{f\}$ we denote the set of all admissible flows, then $\mathcal{G}_0 = \{f_0 \in \mathcal{G} \mid \forall\ f \in \mathcal{G},\ V(f_0) \geqslant V(f)\}$ is the set of all the flows of maximal magnitude.

Let us assume that on the set of all the arcs of the graph we have specified a weighting function φ with values in R, $\varphi: \Gamma \to R$. We define the weighting function $\varphi(f)$ of flow f thus:

$$\varphi(f) = \bigodot_{u \in \Gamma} \psi(\varphi(u), f(u)),$$

where the function $\psi: R \times R^1_+ \to R$ satisfies the following conditions: $\forall a \in R$, $\forall b, c \in R^1_+$, $\psi(a, b + c) = \psi(a, b) \odot \psi(a, c)$;

Table 1.5.3.6

№	Maximal flow	R	$\oplus$	0_R	$\odot$	1_R	$\bar{\varphi}(u)$	$\psi(a, b)$, $b > 0$
1	Minimal cost	R^1	min	$+\infty$	$+$	0	$-\varphi(u)$	$a \times b$
2	Minimax	R^1	min	$+\infty$	max	$-\infty$	$\varphi(u)$	a
3	Maximin	R^1	max	$-\infty$	min	$+\infty$	$\varphi(u)$	a
4	Most reliable	R^1_+	max	0	$\times$	1	$\varphi^{-1}(u)$	a^b

$\forall\, a \in R$, $\psi(a, 0) = 1_R$ (see Table 1.5.3.6). We wish to find the flow $f_0 \in \mathfrak{A}_0$ for which the following condition holds true:

$$\varphi(f_0) = \bigoplus_{f \in \mathfrak{A}_0} \varphi(f).$$

At each ith iteration, $i = 0, 1, 2, \ldots, n(P_0) - 1$, of the algorithm for constructing the optimal f_0 the following sequence of steps is carried out:

(1) by the flow f constructed in the previous iteration i we determine the medium $M^i = (X, \Gamma^i, L^i, R_\oplus)$: to each arc $u \in \Gamma$ in Γ^i there generally correspond two arcs, the "direct" (with the same orientation) arc u with carrying capacity $C(u) - f(u) > 0$ and the "reverse" (with the opposite orientation) arc $\bar{u}$ with carrying capacity $f(u) > 0$; $\forall a \in R$, $L^i(u)(a) = \varphi^i(u) \odot a$, where $\varphi^i(u) = \varphi(u)$ for the direct arc and $\varphi^i(\bar{u}) = \bar{\varphi}(u)$ for the reverse are defined in Table 1.5.3.6;

(2) in medium M^i we solve the Bellman equation corresponding to the problem of finding the optimal path $\mu_0 \in M_{X_1 \to X_2}$ (see (1.5.3.3));

(3) along the established path μ_0 the flow increases by the magnitude of the carrying capacity along this path.

The optimality of the flow algorithm constructed here follows from the fact that at each ith iteration we have for the constructed flow

$$\psi(f) = \bigoplus_{f' \in \mathfrak{A}^i} \psi(f').$$

where $\mathfrak{A}^i = \{f' \in \mathfrak{A} \mid V(f') = V(f)\}$ is the set of all the flows of magnitude $V(f)$.

In Table 1.5.3.6 we list a number of semi-rings corresponding to problems of optimal flows in transport networks.

References

1.1. R. Bellman, *Dynamic Programming* (Princeton, N.J.: Princeton Univ. Press, 1957).

1.2. L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, and E.F. Mishchenko, *The Mathematical Theory of Optimal Processes* (New York: Wiley, 1962).

1.3. A.D. Ioffe and V.M. Tihomirov, *Theory of Extremal Problems* (Amsterdam: North-Holland, 1979).
1.4. S.Y. Kung, K.S. Arun, D.V. Bhaskar Rao, and Y.H. Hu, in: *Proc. CMU Conf. VLSI Syst. Computations* (Comput. Sci. Press, Oct. 1981): pp. 235-244.
1.5. S.Y. Kung, R.J. Gal-Ezer, and K.S. Arun, *Wavefront Array Processor: Architecture, Language, and Applications* (Presented at the M.I.T. Conf. Advanced Res. in VLSI) (Cambridge, Mass.: M.I.T. Press, Jan. 1982).
1.6. U. Weiser and A. Davis, in: *Proc. CMU Conf. VLSI Syst. Computations* (Comput. Sci. Press, Oct. 1981): pp. 226-234.
1.7. S.Y. Kung, K.S. Arun, R.J. Gal-Ezer, and D.V. Bhaskar Rao, *IEEE Trans. Comput.* C-31, No. 11: 1054-1066 (1982).
1.8. J.M. Ziman, *Principles of the Theory of Solids*, (London: Cambridge Univ. Press, 1964).
1.9. L.R. Ford and D.R. Fulkerson, *Flows in Networks* (Princeton, N.J.: Princeton Univ. Press, 1962).
1.10. Yu.L. Ziman and G.G. Ryabov, in: *Electronic Computers* (Moscow: ITM and VT Akad. Nauk SSSR, 1965): pp. 56-110. (in Russian).
1.11. V.V. Belov, E.M. Vorob'ev, and V.E. Shatalov, *The Theory of Graphs* (Moscow: Vysshaya shkola, 1975) (in Russian).
1.12. V.P. Maslov and V.G. Danilov, in: *Proc. Steklov Inst. Math.* **1**: 143-177 (1986); *ibid* **2**: 103-116.
1.13. V.P. Maslov, *Operational Methods* (Moscow: Mir Publishers, 1976).
1.14. M.V. Karasev and V.P. Maslov, in: *Current Problems in Mathematics*, vol. 13 (Moscow: VINITI, 1979): pp. 145-267 (in Russian).
1.15. R.E. Edwards, *Functional Analysis* (New York: Holt, Rinehart, and Winston, 1965).
1.16. Yu.K. Dmitriev and V.G. Khoroshevskii, *Computational Systems of Minicomputers* (Moscow: Radio i svyaz', 1982) (in Russian).
1.17. I.V. Prangishvili, S.Ya. Vilenkin, and I.L. Medvedev, *Parallel Computational Systems with General Control* (Moscow: Energoizdat, 1983) (in Russian).
1.18. A.P. Ershov (ed.), *Algorithms, Software, and Architecture of Multiprocessor Computational Systems* (Moscow: Nauka, 1982) (in Russian).
1.19. E.V. Evreinov, *Homogeneous Computational Systems* (Moscow: Radio i svyaz', 1981) (in Russian).
1.20. B.A. Dubrovin, S.P. Novikov, and A.T. Fomenko, *Modern Geometry: Methods and Applications*, 3 vols. (Berlin: Springer 1984, 1985).
1.21. A.I. Anselm, *Introduction to Semiconductor Theory* (Moscow: Mir Publishers, 1981).
1.22. V.P. Maslov, *Complex Markov Chains and the Continual Feynman Integral for Nonlinear Equations* (Moscow: Nauka, 1976) (in Russian).
1.23. V.V. Belov and V.P. Maslov, in: *The Theory of Graphs* (Moscow: Vysshaya shkola, 1975): pp. 295-389 (in Russian).
1.24. S.M. Avdoshin, V.V. Belov, and V.P. Maslov, *The Discrete Huygens-Fresnel Principle and Its Application in Fifth Generation Computational Systems: Design of Architecture, Parallelism, and Lexical Analysis of EYa and GAP VS*-5 (Deposited in VNTITs, Moscow, No. 02840034978, 1983) (in Russian).
1.25. A.N. Karasev, in: *Problems of Applied Mathematics and Mechanics* (Moscow: Nauka, 1971): pp. 70-95 (in Russian).
1.26. A.P. Ershov, *Kibernetika* (English transl.: Cybernetics) No. 1: 9-20 (1973).
1.27. S.E. Madnick and J.J. Donovan, *Operating Systems* (New York: McGraw-Hill, 1974).
1.28. V.A. Zhirov, *Software and the Design of Computer Structures* (Moscow: Nauka, 1979) (in Russian).

1.29. *Outline of Research and Development Plans for Fifth Generation Computer Systems* (ICOT, May 1982).
1.30. T. Moto-Oka (ed.), *Fifth Generation Computer Systems: Proc. Intl. Conf., Tokyo, Oct. 19-22, 1981* (Amsterdam: North-Holland, 1982).
1.31. G.L. Simons, *Towards Fifth-Generation Computers* (NCC Publications, 1983).
1.32. S.A. Maiorov and G.V. Orlovskii (eds.), *Flexible Automatic Manufacturing* (Leningrad: Mashinostroenie, 1983) (in Russian).
1.33. S.P. Mitrofanov, *Group Technology in Machinery Industry* (Leningrad: Mashinostroenie, 1983) (in Russian).
1.34. M.M. Bongard, *Recognition Problems* (Moscow: Nauka, 1967) (in Russian).
1.35. J. Van Ryzin, *Classification and Clustering: Proc. of Advanced Seminar Conducted by the Mathematical Research Center at University of Wisconsin-Madison* (May 3-5, 1976) (New York: Academic Press, 1977).
1.36. N.G. Zagoruiko, *Recognition Methods and Their Application* (Novosibirsk: Nauka, 1972) (in Russian).
1.37. N. Christofides, *Graph Theory: An Algorithmic Approach* (New York: Academic Press, 1975).
1.38. A. Kaufmann, *Introduction in theorie des sous-ensembles flous* (Paris: Masson, 1977).
1.39. S.M. Avdoshin, N.A. Babaev, and G.Ya. Voloshin, in: *Discovery of Empirical Laws with the Aid of Computers: Computational Systems*, vol. 102 (Novosibirsk: IM SO Akad. Nauk SSSR, 1984): pp. 94-104 (in Russian).
1.40. R. E. Bellman and L.A. Zadeh, *Management Science* **17**, No. 4: 141-164 (1970).
1.41. V.B. Kuz'min, *Constructing Group Solutions in Spaces of Precise and Fuzzy Binary Relations* (Moscow: Nauka, 1982) (in Russian).
1.42. E.E. Beckenbach (ed.), *Applied Combinatorial Mathematics* (New York, Wiley, 1964).
1.43. V.I. Bregman, *Graphs in Problems of Manufacturing Control* (Moscow: Statistika, 1974) (in Russian).
1.44. G.M. Adel'son-Vel'skii, E.A. Dinits, and A.V. Karzanov, *Flow Algorithms* (Moscow: Nauka, 1975) (in Russian).
1.45. R.G. Busacker and T.L. Saaty, *Finite Graphs and Networks* (New York: McGraw-Hill, 1973).
1.46. J. Edmonds and R.M. Karp, *J. Assoc. Comput. Mach.* **19**, No. 2: 248-264 (1972).
1.47. E.A. Dinits, *Dokl. Akad. Nauk SSSR* **194**, No. 4: 754-757 (1970).
1.48. A.V. Karzanov, *Dokl. Akad. Nauk SSSR* **215**, No. 1: 49-52 (1974).
1.49. B.V. Cherkasskii, in: *Mathematical Methods of Solving Economics Problems*, vol. 7 (Moscow: Nauka, 1977): pp. 117-126 (in Russian).
1.50. Z. Galil, in: *Proc. 19th Annual Symp. Foundations of Computer Science* (IEEE Computer Society, 1978): pp. 231-245.
1.51. Z. Galil and A. Naamad, in: *Proc. 11th Annual ACM Symp. Theory of Computing* (ACM, 1979): pp. 13-26.
1.52. D.D. Sleator (Unpublished dissertation. Stanford University).
1.53. *1st Intl. Federation for Information Processing Conf. on Computer Application, on Production and Engineering* (Amsterdam, The Netherlands, Apr. 25-28, 1983): pp. 23-30.
1.54. *2nd Intl. Federation of Automatic Control Symp. on Computer Aided Design of Multivariable Technological Systems* (West Lafayette, Ind., Sept. 15-17, 1982): pp. 101-110.
1.55. E.M. Reingold, J. Nievergelt, and N. Deo, *Combinatorial Algorithms: Theory and Practice* (Englewood Cliffs, N.J.: Prentice-Hall, 1977).
1.56. S.M. Avdoshin and V.V. Belov, *Zh. Vych. Mat. i Mat. Fiziki* **19**, No. 3: 739-755 (1979).

1.57. K.A. Volosov, V.G. Danilov, and V.P. Maslov, *Mathematical Models of Manufacturing Technology of VLSI Circuits* (Moscow: VINITI, 1984) (in Russian).
1.58. V.P. Maslov, *Mathematical Aspects of Integral Optics* (Moscow: VINITI, 1983).
1.59. V.P. Maslov, *Resonance Processes in the Wave Theory and Self-focalization* (Moscow: VINITI, 1983).
1.60. K. Edwards and R. Yasaku, *Datamation* **27**, No. 3: 21-27 (1981).
1.61. *Proc. 8th Intl. Symp. on Industrial Robots*, vol. 1 (Stuttgart, Fed. Rep. of Germany, 1978): pp. 373-384.
1.62. A. Astrop, *Machinery and Production Engineering* **17**: 120-124 (Feb. 1982).
1.63. S. Ando, in: *Proc. 4th Intl. Symp. on Industrial Robots* (Tokyo: 1974): pp. 385-394.
1.64. T. Lozano-Perez, *Proc. IRE* **69**, No. 1 (1981).
1.65. T. Lozano-Perez, *Robotics, Artificial Intelligence, An International Journal*, **19**, N 2, 137-143 (1982).
1.66. J. Hartley, *FMS at Work* (Amsterdam: North-Holland, 1984).
1.67. V.I. Arnol'd, *Classical Mechanics* (Moscow: Nauka, 1974) (in Russian).
1.68. L.V. Kantorovich and G.P. Akilov, *Functional Analysis* (Oxford: Pergamon Press, 1982).
1.69. S.M. Avdoshin, V.V. Belov, and V.P. Maslov, *Mathematical Aspects of Design of Computational Media* (Moscow: VINITI, 1984) (in Russian).
1.70. A.V. Aho, J.E. Hopcroft, and J.D. Ullman, *The Design and Analysis of Computer Algorithms* (Reading, Mass.: Addison-Wesley, 1976).
1.71. G.S. Plesnevich and M.S. Saparov, *Algorithms in the Theory of Graphs* (Ashkhabad: Ylym, 1981) (in Russian).
1.72. K.L. Cooke and E. Halsey, *J. Math. Anal. Appl.* **14**, No. 3: 493-498 (1966).
1.73. C. Berge, *Theorie des graphes et ses applications* (Paris: Dunod, 1958).
1.74. M. Gondran, in: *Combinatorial Programming Methods and Algorithms*, ed. by B.Roy (Dordrecht: E.Reidel Publishing Company, 1975): pp. 137-148.

2 Design of the Optimal Dynamic Analyzer: Mathematical Aspects of Sound and Visual Pattern Recognition

V.P. Belavkin and V.P. Maslov

2.0 A Brief Survey

The problem of automatic recognition of sound signals has in the last several years come to the forefront in connection with one of the most important problems in building fifth generation computational systems, that of voice input and output. Such automatic recognition of sound patterns contains a number of mathematical subproblems, the most important of which is the (mathematical) design of an optimal dynamic analyzer, the device occupying the center of the stage in the problem of automatic recognition of sound patterns. An example of such a device is given in [2.1]: a receiver of acoustic waves $v(x - ct)$ whose idealized model is a point-like resonator (or cavity) capable of measuring the intensities of the vibrational (v) modes excited by the waves. The modes are the natural vibrations of one or several standards placed at point $x = 0$ and are described by an orthonormal set of functions $\chi_k(t)$ on a given interval of observation $[0, T]$. A typical example of such a resonator is the spectrum analyzer, a device that measures the intensity distribution over the discrete frequencies $f_k = k/T$, $k \in N$, and can be represented by a selective filter of harmonic waves

$$v_k(x - ct) = 2\,\mathrm{Re}\,\psi^k \exp\{2\pi jk(x - ct)/cT\}, \qquad j = \sqrt{-1},$$

in the output of which one can measure the positive numbers $\nu^k = |\psi^k|^2$ determined by the complex-valued amplitudes $\psi^k \in \mathbb{C}^1$. The spectrum selector described by the harmonic functions $\chi_k(t) = \exp\{-2\pi jkt/T\}$, $k = 0, 1, \ldots$, which, obviously, form an orthonormal set $\{\chi_k\}$ with respect to the scalar product

$$(\chi_i|\chi_k) = T^{-1}\int_0^T \chi_i^*(t)\,\chi_k(t)\,dt$$

is ideally suited for the discrimination of pure tones with multiple frequencies $\{f_i\}$, tones described by disjoint complex-valued amplitudes $\psi_i^k = 0$ at $i \neq k$ corresponding to the harmonic waves $v_i(x - ct) = 2\,\mathrm{Re}\,\psi_i(t - x/c)$, where $\psi_i(t) = \sum_{k=0}^{\infty}\psi_i^k\chi_k(t)$. To establish

which of the tones in $\{\psi_i\}$ with different frequencies and nonzero intensities $\nu_i = |\psi_i^k|^2 \neq 0$ at $i = k$ is actually detected by such a receiver, it is sufficient to find the number i of the excited standard tuned to one of the harmonic modes in $\{\chi_i\}$ corresponding to the set $\{\psi_i\}$: $\chi_i(t) = \psi_i(t)/\|\psi_i\|$. The vibrational energy of such a standard will coincide with the intensity of the detected signal ψ_i: $|(\chi_i \mid \psi_i)|^2 = \|\psi_i\|^2$, while the other standards remain unexcited: $(\chi_k \mid \psi_i) = 0$ at $k \neq i$. A segment of human speech of duration T containing a finished sentence consists, however, not of a single pure tone but, generally, of an infinitude of pure tones of different amplitudes and frequencies (the frequencies may be assumed to be multiples of $1/T$ at a fixed interval T of a single reception act). Nonharmonic signals, described by spectral amplitudes $\psi_i = [\psi_i^k]_{k=0}^{\infty}$ or, in the temporal representation, by the analytic signals

$$\psi_i(t) = \sum_{k=0}^{\infty} \psi_i^k \exp\{-2\pi jkt/T\},$$

may be indistinguishable in spectral measurements, even if they are orthogonal. For example, if the $\psi_i(t)$, $i = 1, \ldots, m$ are disjoint pulses obtained through the shift by T of the pulsed signal $\psi_0(t)$ of length $\Delta t = T/m$, these pulses have corresponding to them the orthogonal spectral amplitudes $\psi_i^k = \psi_0^k \exp\{2\pi jik/m\}$ with the same intensity distributions $\nu_i^k = |\psi_i^k|^2 = |\psi_0^k|^2$, $i = 1, \ldots, m$, $k \in N$.

Orthogonal sound signals $\{\psi_i\}$ on $[0, T]$ may be identified in a similar manner by selective filters matched with the signal modes $\chi_i(t) = \psi_i(t)/\|\psi_i\|$ that measure the intensity distribution $\nu^k = (\chi_k \mid \psi_i)|^2$ in the modes $\{\chi_i\}$, a distribution that has a different form for different values of i, namely, $\nu_i^k = \|\psi_i\|^2 \delta_i^k$. However, different sentences in human speech correspond ordinarily to nonorthogonal sound signals $\psi_i(t)$, which from the viewpoint of their meaning are identified if they are collinear, that is, differ only in the total energy $\|\psi_i\|^2 = \int |\psi_i(t)|^2\, dt$. For the recognition of nonorthogonal signals $\{\psi_i\}_{i=1}^{m}$ one cannot employ matched filtration since the filters described by the nonorthogonal modes $\chi_i = \psi_i/\|\psi_i\|$ are noncommutative and, hence, cannot be matched in a single selector; otherwise, the total measured intensity $\sum_{k=1}^{m} |(\chi_k \mid \chi_i)|^2$ would exceed the total energy $\|\psi_i\|^2$ of the received signal ψ_i if $(\psi_i \mid \psi_k) \neq 0$ at least for one $k \neq i$. Thus, we have an indefinite situation, formally similar to the incompatibility of noncommutative quantum mechanical observables, a situation arising from Bohr's complementary principle [2.2] and Heisenberg's uncertainty relation [2.3]. Typical

examples of noncommutative filters are the frequency and temporal filters, which are incompatible, just as position and momentum measurements are incompatible in a quantum mechanical system. So which of the disjoint selectors described by orthonormal sets of modes $\{\chi_k\}$ must we employ to discern the nonorthogonal sound signals from a given set $\{\psi_i\}$? It is natural to look for the answer to this nontrivial question in the form of a solution to an optimization problem by selecting a criterion of discernment quality such that the optimal selector does not depend on the gauge transformation $\psi_i \to \lambda\psi_i$ for every complex-valued λ. The latter condition is satisfied by the criterion of the maximum of the total intensity $\sum_{i=1}^{m} |(\chi_i \mid \psi_i)|^2$ of the true received amplitudes or by the criterion of the minimum of the lost intensity $\sum_{i\neq k} |(\chi_i \mid \psi_i)|^2$. The corresponding extremal problem for arbitrary nonorthogonal amplitudes $\{\psi_k\}$ describing quantum mechanical states normalized to a priori probabilities was first studied in general form in [2.4-2.6]. Particular solutions of this problem for the case of two nonorthogonal amplitudes $\{\psi_0, \psi_1\}$ were obtained in [2.7, 2.8], while the case of several linearly independent amplitudes $\{\psi_i\}$ was also considered in [2.9, 2.10].

The above-noted analogy between optimal recognition of sound signals and discernment of quantum mechanical states suggested the possibility of constructing a wave theory of noncommutative measurements within the scope of which one could solve more general problems of testing wave hypotheses for estimating the wave parameters. From the formal viewpoint this theory generalizes the quantum theory of optimal measurements, hypothesis testing, and estimation of parameters [2.11, 2.12], while actually it carries the mathematical methods and ideas developed in [2.4-2.35] into the new, more realistic, field of applications. A short author's review of optimal processing of quantum signals is given in [2.30]. (Additional literature on the quantum theory of detection, hypothesis testing, and parameter estimation can be found in the references cited in [2.11, 2.12].)

In the present paper we give a systematic description of the wave theory of representation and measurement based on analogies with quantum mechanics. This theory is then employed to solve the problems of detection, discrimination, identification, and estimation of the parameters of sound signals and visual patterns within the framework of the noncommutative theory of testing wave hypotheses developed here. The idea of employing the methods of quantum mechanics for discerning wave patterns emerged at the beginning of the 1970s, when a seminar on quantum mechanics and pattern recognition was organized in the Physics Department of Moscow State University. The seminar was directed by one of the authors of

the present paper, V. P. Maslov, along with Yu. P. Pyt'ev, and the second author, V. P. Belavkin, attended it. Interest in problems of registration and reconstruction of the complete wave field rather than only the energy illumination of the image in a certain plane was stimulated by the rapid development of holography, which was invented by Gabor [2.36] and then underwent a revival [2.37] when coherent sources of light, or lasers, were created. The emerging optimization problems of discerning wave fronts are also similar to the problems of discerning quantum mechanical states and cannot be solved by classical methods [2.38] of pattern recognition since it is impossible to register directly and exactly the phase and amplitude of a wave field by measuring the energy parameters. A detailed study of these problems at the time showed [2.39] that the then existing quasiclassical methods of solving quantum mechanical problems were also inadequate, and the solution had to be postponed until a consistent noncommutative theory of measurements was developed in the then rapidly advancing field of quantum theory.

Note that some particular problems of processing optical wave signals, such as those of optical localization [2.40], detection, and discrimination of two signals from closely positioned sources of coherent radiation [2.41, 2.42], have been well studied by methods of quantum statistical and nonlinear optics [2.43-2.45].

In addition to discussing the noncommutative theory of measurements common for sound signals and optical waves, we give solutions to a number of new problems (one of the authors, V. P. Belavkin, obtained these solutions earlier, when solving similar problems for quantum signals). Content and commentaries to the list of literature are presented in brief summaries at the beginning of each section. Similar summaries are given at the beginning of each subsection. The subsections are written as a series of articles of increasing complexity so that each can be read independently, although the best way to understand the material is to carefully read the articles in the order given.

2.1 Representation and Measurement of Acoustic Signals and Optical Fields

In this section we discuss the mathematical apparatus of the wave theory of representation and measurement of sound and visual patterns. Along with pure wave patterns, which are described by coherent signals and fields, we also consider the representation of mixed patterns, which are described by partially coherent and incoherent signals and fields. In addition to the spatial-frequency (coordinate) and wave-temporal (momentum) representations we introduce the joint canonical representation, in which the pure and

mixed patterns are described by entire functions and kernels in a phase complex space. We develop the mathematical theory of ideal filters and quasifilters, disjoint selectors, and quasiselectors, which describe ordinary, successive, and indirect measurements of wave-pattern intensity distributions. Using this theory as a basis, we analyze coordinate and momenta measurements as well as quasi-measurements of joint coordinate-momentum distributions. The mathematical tools used here are in many respects similar to those used in the quantum theory of representations and measurements [2.3], which recently received a new impetus in connection with problems of quantum states recognition [2.4-2.12]; however, we give a wave rather than a statistical interpretation of the apparatus in accordance with the applications considered here.

2.1.1 Mathematical Description of Sound and Visual Patterns

In this section we describe three basic types of representation of pure and mixed wave patterns: the coordinate, or spatial-frequency, the momentum, or wave-temporal, and the canonical, in which the wave patterns are represented by holomorphic amplitudes in the complex coordinate-momentum plane. The third representation, which emerged in quantum optics [2.43], proves useful in an analysis of the frequency-temporal structure of sound and visual patterns and in holography in an analysis of the spatial-temporal structure of such patterns.

2.1.1.1 Wave Patterns

The sound and visual patterns considered here are commonly described by wave amplitudes $v(t, \mathbf{q})$ that are real-valued functions of time t and coordinates $\mathbf{q}$ in a spatial-temporal region accessible for measurement. Although the simplest wave equations describing the behavior of such physical fields are linear in amplitudes $v(t, \mathbf{q})$, for purposes of measurement of sound and visual patterns the most interesting are functionals quadratic in v that describe the distribution of sound on the standards of the dynamic analyzer or of light on photodetectors, a distribution that in spatial-temporal measurements is described by the intensity function $v^2(t, \mathbf{q})/2$. More useful information is provided not by the intensities of a sound or visual pattern at points $(t, \mathbf{q})$ but by the distribution of the sound intensity at frequency f (a spatial-frequency distribution), as is common in an analysis of color patterns. Such a distribution is determined by the intensity function

$$i(f, \mathbf{q}) = |\varphi(f, \mathbf{q})|^2, \quad f \geqslant 0, \tag{2.1.1.1}$$

with φ the complex-valued spectral amplitudes,

$$\varphi(f, \mathbf{q}) = \int_{-\infty}^{+\infty} v(t, \mathbf{q})\, e^{2\pi j f t}\, dt,$$

used to represent the wave field $v(t, \mathbf{q})$ in the form of a linear combination of harmonic oscillations at each point $\mathbf{q}$:

$$v(t, \mathbf{q}) = 2\,\mathrm{Re}\int_0^\infty \varphi(f, \mathbf{q})\, e^{-2\pi j f t}\, df.$$

Bearing in mind the well-known advantage of employing complex-valued amplitudes, in what follows we consider complex-valued signals $\varphi(x) = \varphi(f, \mathbf{q}) \equiv \varphi(q)$, with $x = (f, \mathbf{q}) \equiv q$, assuming that in a given spatial-frequency region of measurements Ω they possess a finite total intensity

$$I(\varphi) = \int_\Omega |\varphi(q)|^2\, dq. \tag{2.1.1.2}$$

When registering sound signals for which the spatial region of measurements is much smaller than the characteristic length of the sound wave, we can take for Ω a one-dimensional region, which is usually determined by a positive-frequency pass band $\Phi \in \mathbb{R}_+$ of the dynamic analyzer, assuming that $x = f$ at the point of its localization $\mathbf{q} = 0$; recognizing visual patterns usually requires only a three-dimensional region $\Omega = \Phi \times S$, where Φ is the optical frequency band and S the surface on which the pattern is localized; for static patterns $x = \mathbf{q}$ $(t = 0)$.

From the standpoint of physics the admissible amplitudes are smooth amplitudes $\varphi(q)$, $q \in \Omega$, with compact supports inside a $(d + 1)$-dimensional region $\Omega \subseteq \mathbb{R}^{d+1}$ or, if Ω is noncompact, amplitudes that fall off rapidly at infinity together with all their derivatives. Such amplitudes generate a Hilbert space $\mathcal{H} = \mathcal{L}^2(\Omega)$ of amplitudes $\chi(q)$ of finite intensity $\|\chi\|^2 < \infty$:

$$(\varphi|\chi) = \int_\Omega \varphi^*(q)\,\chi(q)\, dq. \tag{2.1.1.3}$$

Generally the set $\mathcal{D}$ of basic amplitudes φ can form an arbitrary complex-valued space with a positive Hermitian form $I(\varphi) = (\varphi \mid \varphi)$ that is invariant with respect to complex conjugation $\varphi \mapsto \varphi^*$. This form defines a finite intensity, $I(\varphi) \neq 0$, for any nonzero φ. We denote the completion of this space in norm $\|\varphi\| = I^{1/2}(\varphi)$ by $\mathcal{H}$ and consider it to be a Hilbert space equipped with an isometric involution $\chi \mapsto \chi^*$ with respect to the scalar product (2.1.1.3), which is linear in the second argument $\varphi \in \mathcal{H}$.

The use of complex-valued amplitudes not only considerably simplifies the formulas for calculating the observed distributions of the fields but also makes it possible to employ analogies from quantum theory. Specifically, as in quantum theory, complex-valued amplitudes differing in a phase factor must be assumed "equal" since they

lead to the same intensities defined by Hermitian forms of φ. Note that in the quantum description of optical and sound signals we usually take the mean number of the corresponding quanta (photons and phonons) as the intensity functions for light and sound, respectively. These quantities are determined by the same Hermitian forms of the complex-valued amplitudes φ as in the classical mode of description, provided that the quantum mechanical states are coherent [2.43, 2.44], that is, are described by Poisson probability amplitudes $|\varphi\rangle$. Thus, restricting ourselves to intensity measurements, we postulate that only distributions of quanta are observable, while the only characteristics of signals φ of interest to physics are those obtained as a result of measurement of such distributions.

2.1.1.2 Momentum Representation

In problems dealing with the recognition of moving patterns what may be of interest is not the spatial-frequency intensity distribution (2.1.1.1) but the momentum-temporal distribution described by the function

$$\tilde{\iota}(t, \mathbf{p}) = |\tilde{\varphi}(t, \mathbf{p})|^2, \quad (t, \mathbf{p}) \in \mathbb{R}^{d+1}, \tag{2.1.1.4}$$

where $\tilde{\varphi}(t, \mathbf{p})$ is the involution Fourier transform,

$$\tilde{\varphi}(t, \mathbf{p}) = \int \int \varphi^*(f, \mathbf{q}) (e^{2\pi j(tf + \mathbf{p}\cdot\mathbf{q})} df\, d\mathbf{q}. \tag{2.1.1.5}$$

We introduce the notation $x = (t, \mathbf{p}) \equiv p$. The representation of amplitudes φ in terms of the functions $\tilde{\varphi}(x) = \tilde{\varphi}(t, \mathbf{p}) \equiv \tilde{\varphi}(p)$ is called the momentum representation.[1] Note that this representation differs from the common one by complex conjugation, $\widetilde{\varphi^*} = \tilde{\varphi}^*$, but this difference is "unobservable" from the viewpoint of measuring intensities described by Hermitian forms (2.1.1.1) and (2.1.1.4), which are invariant under such conjugation.

Allowing for Plancherel's equality

$$\int |\tilde{\varphi}(p)|^2 dp = \int |\varphi(q)|^2 dq, \tag{2.1.1.6}$$

we find that the total intensity described by the distribution function (2.1.1.4) coincides with the total intensity (2.1.1.2) for any amplitude

---

[1] Thanks to the introduction of the involution $\varphi \mapsto \varphi^*$ in transformation (2.1.1.5), the inverse transformation to the coordinate representation is carried out by the same formula (2.1.1.5), $\varphi = \tilde{\tilde{\varphi}}$, which can be extended onto generalized amplitudes χ in the standard manner.

φ with support in Ω:

$$\int \tilde{\iota}(p)\,dp = I(\varphi) = \int \iota(q)\,dq. \tag{2.1.1.7}$$

In what follows we consider the values of $\iota(q)$ and $\tilde{\iota}(p)$ in (2.1.1.1) and (2.1.1.4) as being functions of φ, denoting by $\iota(\varphi, q)$ and $\tilde{\iota}(\varphi, p)$ the functionals connected by involutions $\tilde{\iota}(\varphi, p) = \iota(\tilde{\varphi}, p)$ and $\tilde{\tilde{\iota}}(\varphi, q) = \iota(\varphi, q)$, respectively. Note that the Fourier transformation $\iota(q) \mapsto \tilde{\iota}(p)$ of the Hermitian functional $\iota(\varphi, q) = |\varphi(q)|^2$ cannot be reduced to the Fourier transformation of its value $\iota(\varphi)$ as a function of q. More than that, measuring the spatial distribution $\tilde{\iota}(q)$ for a single value of φ does not generally make it possible in any way to calculate the corresponding distribution $\tilde{\iota}(p)$, and vice versa. Nevertheless, these distributions satisfy certain relationships, the simplest of which are (2.1.1.7) and the inequality

$$\int (p-\bar{p})^2\,\tilde{\iota}(p)\,dp \int (q-\bar{q})^2\iota(q)\,dq \geqslant \frac{1}{(4\pi)^2}\,I^2(\varphi) \tag{2.1.1.8}$$

(where $\bar{p} = \int p\tilde{\iota}(p)\,dp/I(\varphi)$ and $\bar{q} = \int q\iota(q)\,dq/I(\varphi)$ for each of the components $p = p_k$ and $q = q_k$, $k = 0, \ldots, d$), which is known as the uncertainty relation. Using the commutation relations

$$\hat{p}\hat{q} - \hat{q}\hat{p} = (2\pi j)^{-1}\hat{1} \tag{2.1.1.9}$$

for each pair of operators $\hat{q}$, $\hat{p}$ in the p-representation, with $\hat{p} = p - \bar{p}$ and $\hat{q} = \partial/\partial(2\pi j p) - \bar{q}$, we can easily arrive at (2.1.1.8) as a corollary of Schwarz's inequality

$$\|\hat{p}\tilde{\varphi}\|\,\|\hat{q}\tilde{\varphi}\| \geqslant |(\hat{p}\tilde{\varphi}|\hat{q}\tilde{\varphi})| \geqslant |\mathrm{Im}\,(\hat{p}\tilde{\varphi}|\hat{q}\tilde{\varphi})| = \frac{1}{2\pi}\,\|\tilde{\varphi}\|^2. \tag{2.1.1.10}$$

Indeed, according to definition (2.1.1.4) we have

$$\int (p-\bar{p})^2\,\tilde{\iota}(p)\,dp = \int |(p-\bar{p})\,\tilde{\varphi}(p)|^2\,dp = \|\hat{p}\tilde{\varphi}\|^2$$

and, similarly, allowing for (2.1.1.6), for the Fourier transform $\hat{p}\tilde{\varphi}$ of the function $(q-\bar{q})\,\varphi(q)$ we obtain from (2.1.1.1)

$$\int (q-\bar{q})^2\,\iota(q)\,dq = \int |(q-\bar{q})\,\varphi(q)|^2\,dq = \|\hat{q}\tilde{\varphi}\|^2.$$

Thus, inequality (2.1.1.8) is equivalent to (2.1.1.10), where $\|\tilde{\varphi}\|^2 = I(\varphi)$. For nonzero amplitudes φ this inequality is usually written as

$$\sigma_p\sigma_q \geqslant 1/4\pi, \tag{2.1.1.11}$$

where σ_p and σ_q are the standard deviations,

$$\sigma_p^2 = \int (p-\bar{p})^2 \tilde{\iota}(p)\, dp/I(\varphi),$$

$$\sigma_q^2 = \int (q-\bar{q})^2 \iota(q)\, dq/I(\varphi), \qquad (2.1.1.12)$$

of momentum p and coordinate q in the wave packet φ from their mean values $\bar{p}$ and $\bar{q}$. In this form (2.1.1.11) is similar to the quantum mechanical Heisenberg uncertainty relation; however, here the standard deviations (2.1.1.12) have no statistical meaning but characterize the extent to which the intensity distributions are localized in the coordinate and momentum spaces. The lower bound $1/(4\pi)$ in this relation is achieved only in the case of an unbounded region $\Omega = \mathbb{R}^{d+1}$ for the Gaussian amplitudes

$$\varphi(q) = C_q \exp\{2\pi j (q-\bar{q}/2)\bar{p} - |q-\bar{q}|^2/(4\sigma_q^2)\}. \qquad (2.1.1.13)$$

These amplitudes, with the normalization constants $C_q = 1/(2\pi\sigma_p^2)^{(d+1)/4}$, have a similar form in the p-representation:

$$\tilde{\varphi}(p) = C_p \exp\{2\pi j (p-\bar{p}/2)\bar{q} - |p-\bar{p}|^2/(4\sigma_p^2)\},$$

with $C_p = 1/(2\pi\sigma_p^2)^{(d+1)/4}$ and $\sigma_p\sigma_q = 1/(4\pi)$, are called standard canonical (Poisson) amplitudes and are denoted by $\psi_\alpha = |\alpha)$, with $\alpha = (1/2)(\bar{q}/\sigma_q + j\bar{p}/\sigma_p)$ if σ_q and σ_p are fixed. Note that for $\alpha \neq \alpha'$ such amplitudes are nonorthogonal:

$$(\alpha \mid \alpha') = \exp\{-(1/2)|\alpha'|^2 + \alpha'\alpha^+ - (1/2)|\alpha|^2\}, \qquad (2.1.1.14)$$

with $\alpha'\alpha^+$ the scalar product $\sum_{i=0}^{d} \alpha_i^*\alpha_i'$.

2.1.1.3 Mixed Signals

Due to limits in present-day technology, only a fraction of the information on the intensity distribution of amplitude φ in this or another region can usually be obtained when analyzing sound and visual patterns. For instance, in sound pattern recognition the common method is to use only the frequency or temporal distribution obtained through integration

$$\iota(t) = \int |\varphi(t, \mathbf{q})|^2\, d\mathbf{q}, \quad \tilde{\iota}(t) = \int |\tilde{\varphi}(t, \mathbf{p})|^2\, d\mathbf{p} \qquad (2.1.1.15)$$

of distributions (2.1.1.1) and (2.1.1.4) over the spatial or wave region of measurement. In visual pattern recognition often only black and white patterns are considered. These are obtained as the result of mixing

$$\iota(\mathbf{q}) = \int |\varphi(t, \mathbf{q})|^2\, dt, \quad \tilde{\iota}(\mathbf{p}) = \int |\tilde{\varphi}(t, \mathbf{p})|^2\, dt \qquad (2.1.1.16)$$

of the appropriate color patterns in the spatial or wave region of measurement. To obtain such incomplete distributions there is no need to provide a total description of the signal by amplitude $\varphi(f, \mathbf{q})$. For instance, in describing sound it is sufficient to specify only the Hermitian kernel

$$S(f', f) = \int \varphi(f', \mathbf{q})\,\varphi^*(f, \mathbf{q})\,dq,$$

for which $\iota(f) = S(f, f)$ and $\tilde{\iota}(t) = \tilde{S}(t, t)$, where

$$\tilde{S}(t', t) = \int e^{2\pi j(ft' - f't)} S(f', f)\,df'\,df.$$

Monochrome patterns are defined by a similar kernel $S(\mathbf{q}', \mathbf{q})$, with $\iota(\mathbf{q}) = S(\mathbf{q}, \mathbf{q})$ and $\tilde{\iota}(\mathbf{p}) = \tilde{S}(\mathbf{p}, \mathbf{p})$. Having in mind the possibility of such mixing, we will describe signals in an abridged manner by nonnegative definite operators of intensity density, S, with kernels $S(q, q')$ that have a nonzero trace

$$\operatorname{Tr} S = \int_{\Omega} S(q, q)\,dq \equiv \langle S, I\rangle, \qquad (2.1.1.17)$$

which determines the total intensity, $\iota(S) = \langle S, I\rangle$, of the signal in Ω. To each amplitude $\psi(q)$ we assign a one-dimensional operator $S = |\psi)(\psi|$ with a kernel

$$S(q', q) = \psi(q')\,\psi^*(q), \qquad (2.1.1.18)$$

which defines the amplitude $\psi(q)$ to within a nonessential phase factor $e^{j\theta}$. The diagonal values $\iota(q) = S(q, q)$ describe the coordinate distribution of the intensity of such a signal, while the momentum distribution is described by the diagonal values $\tilde{\iota}(p) = \tilde{S}(p, p)$ of the involution Fourier transform

$$\tilde{S}(p', p) = \int e^{2\pi j(p'q^T - q'p^T)} S(q', q)\,dq'\,dq. \qquad (2.1.1.19)$$

Each such kernel can be obtained as a result of mixing

$$S(q', q) = \int \psi_\alpha(q')\,\psi_\alpha^*(q)\,\nu(d\alpha) \qquad (2.1.1.20)$$

of the one-dimensional kernels corresponding to the normalized amplitudes $\{\psi_\alpha\}$, $\|\psi_\alpha\| = 1$, parameterized by a space A with a nonzero positive measure ν whose mass determines the total intensity $\langle S, I\rangle = \nu(A)$. For example, the kernel $S(\mathbf{q}', \mathbf{q})$ corresponding to a monochrome pattern generated by amplitude $\varphi(f, \mathbf{q})$ can be written in the form (2.1.1.20) for $\psi_f(\mathbf{q}) = \varphi(f, \mathbf{q})/\iota^{1/2}(f)$ on the set A of frequencies Φ equipped with a nonzero measure $\nu(df) = \iota(f)\,df$.

In the case of an arbitrary Hilbert space $\mathscr{H}$, mixed signals are described by density operators S obtained as a result of weak integration

$$S = \int |\psi_\alpha)(\psi_\alpha| \nu(d\alpha) \tag{2.1.1.21}$$

of one-dimensional density operators $S_\psi = |\psi)(\psi|$,

$$|\psi)(\psi|: \chi \in \mathscr{H} \mapsto \psi(\psi \mid \chi) = (\psi \mid \chi)\psi, \tag{2.1.1.22}$$

corresponding to the normalized values $\psi_\alpha \in \mathscr{H}$, $\alpha \in A$, of the vector function $\alpha \mapsto \psi_\alpha$. The operators (2.1.1.21) are kernel-positive and have a finite trace $\mathrm{Tr}\, S = \nu(A)$, with each operator $S: \mathscr{H} \mapsto \mathscr{H}$ being represented in the form (2.1.1.21) via, say, the spectral decomposition

$$S = \sum_i |\psi_i)(\psi_i| \nu_i. \tag{2.1.1.23}$$

Here $\{\psi_i\}$ is the maximal orthogonal set of normalized eigenvectors $\psi_i \in \mathscr{H}$ corresponding to zero eigenvalues ν_i: $S\psi_i = \nu_i\psi_i$, which determine the trace $\mathrm{Tr}\, S = \sum_i \nu_i$.

2.1.1.4 Gaussian Signals

As an example let us consider the important class of mixed canonical signals defined by the integration

$$S = \int |\alpha)(\alpha| \nu(d\xi\, d\eta), \quad \alpha \in \mathbb{C}^{d+1}, \tag{2.1.1.24}$$

of canonical projectors corresponding to the amplitudes $\psi_{\xi\eta} = |\alpha)$, which in the coordinate representation have the following general Gaussian form (cf. (2.1.1.13))

$$\psi_{\xi\eta}(q) = C \exp\left\{2\pi j \left(q - \frac{1}{2}\xi\right)\eta^T - \frac{1}{2}(q-\xi)\,\omega\,(q-\xi)^T\right\}. \tag{2.1.1.25}$$

Here $\omega = \omega^T$ is a symmetric complex-valued $(d+1)$-by-$(d+1)$ matrix with a positive definite real part $\omega + \omega^* = 2\pi(v^+ v)^{-1}$, $\xi = \sqrt{2/\pi}\,\mathrm{Re}\,\alpha v$ and $\eta = \sqrt{2/\pi}\,\mathrm{Re}\, j\alpha^*\tilde{v}$, with $\tilde{v} = v^*\omega/(2\pi)$, are $(d+1)$-dimensional rows, ξ^T and η^T are the corresponding columns, $|C|^2 = \det(v^+ v)^{-1} = |v|^{-2}$ is the normalization constant, and $v^{*T} = v^+ = v^{T*}$ is the Hermitian conjugate of matrix v.

Let $\zeta = (\xi, \eta)$ be a $2(d+1)$-dimensional row and $\nu(d\zeta) = \nu(d\xi\, d\eta)$ a Gaussian measure on $A = \mathbb{R}^{2(d+1)}$ normalized to a certain number $J < \infty$ (the Gaussian intensity) and described (for

J positive) by the following moments:

$$\bar{\xi} = J^{-1} \int \xi \nu (d\zeta) = \lambda, \quad \bar{\eta} = J^{-1} \int \eta \nu (d\zeta) = \varkappa, \qquad (2.1.1.26)$$

$$\begin{bmatrix} \overline{\xi^T \xi} & \overline{\xi^T \eta} \\ \overline{\eta^T \xi} & \overline{\eta^T \eta} \end{bmatrix} = J^{-1} \int \zeta^T \zeta \nu (d\zeta) = \begin{bmatrix} \lambda^T \lambda & \lambda^T \varkappa \\ \varkappa^T \lambda & \varkappa^T \varkappa \end{bmatrix} + \sigma_{\zeta\zeta}, \quad (2.1.1.27)$$

where $\sigma_{\zeta\zeta} = \begin{bmatrix} \sigma_{\xi\xi} & \sigma_{\xi\eta} \\ \sigma_{\eta\xi} & \sigma_{\eta\eta} \end{bmatrix}$ is a nonnegative definite $2(d+1)$-by-$2(d+1)$ matrix. The signals that correspond to such a density operator (2.1.1.24) are characterized by the following first moments:

$$\bar{q} = J^{-1} \int Q(\psi_\zeta) \nu (d\zeta) = \lambda, \quad \bar{p} = J^{-1} \int P(\psi_\zeta) \nu (d\zeta) = \varkappa, \qquad (2.1.1.28)$$

$$\overline{q^T q} = J^{-1} \int (Q^T \psi_\zeta \mid Q \psi_\zeta) \nu (d\zeta) = \overline{\xi^T \xi} + (\omega + \omega^*)^{-1},$$

$$\overline{q^T p} = J^{-1} \int (Q^T \psi_\zeta \mid P \psi_\zeta) \nu (d\zeta) = \overline{\xi^T \eta} + j v^T \tilde{v}/(2\pi), \qquad (2.1.1.29)$$

$$\overline{p^T q} = J^{-1} \int (P^T \psi_\zeta \mid Q \psi_\zeta) \nu (d\zeta) = \overline{\eta^T \xi} - j \tilde{v}^+ v^*/(2\pi),$$

$$\overline{p^T p} = J^{-1} \int (P^T \psi_\zeta \mid P \psi_\zeta) \nu (d\zeta) = \overline{\eta^T \eta} + (\tilde{\omega} + \tilde{\omega}^*)^{-1},$$

where Q and P are the rows of operators of position Q_k and momentum P_k defined in the q-representation via multiplication by q_k and differentiation with respect to q_k, or $(2\pi j)^{-1} \partial/\partial q_k$, $\overline{\omega}/(2\pi) = 2\pi/\omega^*$ and we have allowed for the fact that

$$Q(\psi_\zeta) = (\psi_\zeta \mid Q \psi_\zeta) = \xi, \quad P(\psi_\zeta) = (\psi_\zeta \mid P \psi_\zeta) = \eta, \quad (2.1.1.30)$$

$$\begin{bmatrix} (\hat{q}^T \psi_\zeta \mid \hat{q} \psi_\zeta) & (\hat{q}^T \psi_\zeta \mid \hat{p} \psi_\zeta) \\ (\hat{p}^T \psi_\zeta \mid \hat{q}^T \psi_\zeta) & (\hat{p}^T \psi_\zeta \mid \hat{p} \psi_\zeta) \end{bmatrix} = (v^*, \ j\tilde{v})^+ (v^*, \ \tilde{v})/(2\pi)$$

for the "shifted" operators $\hat{q} = Q - \bar{q} I$ and $\hat{p} = P - \bar{p} I$.

It can easily be demonstrated that for a nonsingular Gaussian measure described by density $n(\zeta) = \nu(d\zeta)/d\zeta$ of the form

$$n(\zeta) = C \exp \left\{ -\frac{1}{2} (\zeta - \theta) \sigma_{\zeta\zeta}^{-1} (\zeta - \theta)^T \right\} \qquad (2.1.1.31)$$

(with $\theta = (\varkappa, \lambda)$ and $C = J/\sqrt{\det 2\pi \sigma_{\zeta\zeta}}$) corresponding to a nonsingular correlation matrix $\sigma_{\zeta\zeta}$ we can select representation (2.1.1.24) of the density operator S by appropriate choice of matrix ω in such a manner that the representation will be described by density

(2.1.1.30) with the matrix $\sigma_{\zeta\zeta}$ of the form

$$\sigma_{\zeta\zeta} = \mathrm{Re}\,(v^*,\ j\tilde{v}) + s\,(v^*,\ j\tilde{v})/\pi \tag{2.1.1.32}$$

where s is a complex-valued positive definite $(d+1)$-by-$(d+1)$ matrix. At this point it is expedient to introduce a complex-valued normal representation characterized by the transition made from $2\,(d+1)$ real variables $\zeta = (\xi,\ \eta)$ to a $(d+1)$-dimensional complex variables $\alpha = (\xi\omega + 2\pi j\eta)\,\gamma^{-1}$, where $\gamma = (\omega + \omega^*)^{1/2}$ in terms of which the density (2.1.1.30) combined with (2.1.1.32) can be written in the following form:

$$n\,(\alpha,\ \alpha^*) = C\,\exp\,\{-(\alpha-\theta)^*\,s^{-1}\,(\alpha-\theta)^{\mathrm{T}}\}, \tag{2.1.1.33}$$

where $\theta = (\varkappa\omega + 2\pi j\lambda)\,\gamma^{-1}$, and $C = J/\det s$ if density (2.1.1.33) is normalized to J with respect to $d\alpha\, d\alpha^* = d\xi\, d\eta$. Note that, as in the case with (2.1.1.13), amplitudes (2.1.1.25) must be written in the form

$$|\,\alpha)\,(q) = (\gamma/\sqrt{2\pi})^{1/2}\exp\left\{(q\gamma - \mathrm{Re}\,\alpha)\,\alpha^{\mathrm{T}} - \frac{1}{2}\,q\omega\,q^{\mathrm{T}}\right\}, \tag{2.1.1.34}$$

with the scalar product defined in (2.1.1.14). However, in contrast to (2.1.1.13), these amplitudes do not realize at $\mathrm{Im}\,\omega \neq 0$ the lower bound in the uncertainty relation (2.1.1.11) while they do realize a more exact lower bound defined by the matrix inequality

$$\det\begin{bmatrix}\sigma_{qq} & \sigma_{qp}\\ \sigma_{pq} & \sigma_{pp}\end{bmatrix} \geqslant 0, \text{ or } \sigma_{pp} \geqslant \sigma_{pq}\sigma_{qq}^{-1}\sigma_{qp}, \tag{2.1.1.35}$$

provided that matrix σ_{qq} is nonsingular. Here

$$\begin{aligned}&\sigma_{qq} = \sigma_{\xi\xi} + \gamma^{-2},\ \sigma_{qp} = \sigma_{\xi\eta} - \omega\gamma^{-2}/(2\pi j),\\ &\sigma_{pq} = \sigma_{\eta\xi} + \omega^*\gamma^{-2}/(2\pi j),\ \sigma_{pp}\\ &\quad = \sigma_{\eta\eta} + \omega^*\gamma^{-2}\omega/(2\pi)^2\end{aligned} \tag{2.1.1.36}$$

are the elements of the correlation matrices,

$$\begin{aligned}\sigma_{qq} &= J^{-1}\int (\hat{q}^{\mathrm{T}}\psi_\alpha\,|\,\hat{q}\psi_\alpha)\,\nu\,(d\alpha),\\ \sigma_{qp} &= J^{-1}\int (q^{\mathrm{T}}\psi_\alpha\,|\,\hat{p}\psi_\alpha)\,\nu\,(d\alpha),\\ \sigma_{pq} &= J^{-1}\int (\hat{p}^{\mathrm{T}}\psi_\alpha\,|\,\hat{q}\psi_\alpha)\,\nu\,(d\alpha),\\ \sigma_{pp} &= J^{-1}\int (\hat{p}^{\mathrm{T}}\,\psi_\alpha\,|\,\hat{p}\psi_\alpha)\,\nu\,(d\alpha),\end{aligned} \tag{2.1.1.37}$$

which are defined for nonmixed canonical signals in (2.1.1.30) and satisfy, obviously, the following relation:

$$\sigma_q(\sigma_p^2 - \rho^* \sigma_p^2 \rho^{\mathrm{T}})\sigma_q = 1/(4\pi)^2, \tag{2.1.1.38}$$

with $\sigma_p^2 = \sigma_{pp}$, $\sigma_q^2 = \sigma_{qq}$, and $\rho = \omega^{-1} \operatorname{Im} \omega$; at $\rho = 0$ this relation realizes the bound of (2.1.1.11). Otherwise, it realizes the bound to (2.1.1.35).

2.1.1.5 Canonical Representations

In problems dealing with the recognition of complicated sound and visual patterns, the most important information is usually contained in the momentum representation as well as in the coordinate representation. For example, in analyzing speech not only the frequency distribution of its intensity is important but so is its temporal distribution, in the same way as in color pattern recognition it has proved important to know the wave structure in addition to the spatial structure. Although there can be no joint coordinate-momentum representation that would enable calculating such distributions simultaneously (due to noncommutativity of position and momentum operators), the simultaneous estimate, say by the human ear, of the frequency and temporal structures of sound points to the possibility of building a mathematical model of such perception, which may prove extremely important for automatic speech recognition.

The simplest models of such joint coordinate-momentum representations are those whose densities are defined as the intensities

$$k(z) = |(\psi_z | \varphi)|^2, \quad z = (x, y) \in \mathbb{R}^{2(d+1)}, \tag{2.1.1.39}$$

of projections of amplitude φ on the canonical amplitudes (2.1.1.24) at $\zeta = z$, which are parametrized at a fixed ω by the estimate vectors $\xi = x$ and $\eta = y$ of the generalized coordinates $x = (x_0, \ldots, x_d)$ and momenta $y = (y_0, \ldots, y_d)$. We can directly verify that the density in x has the form

$$m(x) = \int k(x, y)\, dy = |\upsilon|^{-1/2} \int e^{-\pi |(x-q)\upsilon^{-1}|^2} |\varphi(q)|^2 dq, \tag{2.1.1.40}$$

which in the limit of $|\upsilon| = \sqrt{\det \upsilon^+ \upsilon} \to \infty$ coincides with the coordinate distribution $\iota(q) = |\varphi(q)|^2$. Similarly, in the p-representation we find the density in y:

$$\begin{aligned} \tilde{m}(y) &= \int k(x, y)\, dx \\ &= |\tilde{\upsilon}|^{-1/2} \int e^{-\pi |(y-p)\tilde{\upsilon}^{-1}|^2} |\tilde{\varphi}(p)|^2 dp. \end{aligned} \tag{2.1.1.41}$$

which in the limit of $|\tilde{v}| \to \infty$ coincides with the momentum distribution $\iota(p) = |\tilde{\varphi}(p)|^2$. Here for every amplitude φ we have

$$\|\varphi\|^2 = \int m(x)\,dx = \int\int k(x, y)\,dx\,dy$$
$$= \int \tilde{m}(y)\,dy = \|\tilde{\varphi}\|^2, \qquad (2.1.1.42)$$

which means that the set of canonical amplitudes $\{\psi_z \mid z \in \mathbb{R}^{2d}\}$ is complete for every ω, with Re ω positive. Thus, the $2(d+1)$-parametric set $\{\psi_z\}$ forms a nonorthogonal base that defines for each ω a canonical representation in which the diagonal elements of the kernel $(\psi_z \mid S\psi_z)$ of a signal described by density operator S yield the density of the distribution of the signal's intensity $\iota(S) = \langle S, I\rangle$. For mixed canonical signals (2.1.1.24) such a density with allowance made for (2.1.1.14) has the form

$$k(x, y) = \int \exp\{-|c-\alpha|^2\}\, n(\xi, \eta)\, d\xi, d\eta, \qquad (2.1.1.43)$$

where $c = (x\omega + 2\pi j y)\, v^+/2\pi$, specifically, for Gaussian signals (2.1.1.25) we arrive at the following Gaussian density:

$$k(z) = J \exp\left\{-\frac{1}{2}(z-\theta)\,\sigma_{zz}^{-1}(z-\theta)^{\mathrm{T}}\right\}\Big/\sqrt{\det 2\pi\sigma_{zz}}, \qquad (2.1.1.44)$$

where

$$\sigma_{zz} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} = \sigma_{\zeta\zeta} + 2(v^+, j\tilde{v})^+(v^*, j\tilde{v})/\pi \qquad (2.1.1.45)$$

is a $(d+1)$-by-$(d+1)$ correlation matrix ($j = \sqrt{-1}$).

A remarkable property of canonical representations is the possibility of calculating the intensity distribution in any representation knowing only one canonical distribution by analytically continuing the density $k(c, c^*) = k(x, y)$ to a kernel

$$k(c', c^*) = (c \mid S \mid c') \qquad (2.1.1.46)$$

that is holomorphic in c' and $c^* \in \mathbb{C}^{d+1}$, where $|c) = \psi_z$ and similarly, $|c') = \psi_{z'}$ at $c' = (x'\omega + 2\pi j y')\, v^+/\sqrt{2\pi}$ for $z' = (x', y')$. For one thing, for operator (2.1.1.24) with a density of the form (2.1.1.33) we obtain

$$k(c', c^*) = J(c \mid c') e^{-(c^*-\theta^*)(s+1)^{-1}(c'-\theta')^{\mathrm{T}}}/|S_{\cdot}|. \qquad (2.1.1.47)$$

The transition from kernels of the (2.1.1.46) type to, say, the coordinate representation of operator S is carried out by the following formula:

$$S = \int |c)(c' \mid k(c', c^*)(c \mid c')\, dc\, dc^*\, dc'\, dc^{*\prime}, \qquad (2.1.1.48)$$

where $|c)(c'|: \varphi \mapsto (c'|\varphi)|c)$ are one-dimensional operators defined by the canonical amplitudes (2.1.1.24), respectively, at $\alpha = c$ and $c' \in \mathbb{C}^{d+1}$, $dc\,dc^* = dx\,dy$, $dc'\,dc^{*\prime} = dx'\,dy'$,

$$(c \mid c') = \exp\left\{-\frac{1}{2}\,|c|^2 + c^*c' - \frac{1}{2}\,|c'|^2\right\}. \quad (2.1.1.49)$$

Note that the canonical kernels (2.1.1.46), as operators S, act in the space of entire functions

$$h(c) = e^{|c|^2/2}(\varphi \mid c), \quad (2.1.1.50)$$

which define the representation of amplitudes $\chi \in \mathscr{H}$ in the Bargman space of all entire functions on $\mathbb{C}^{d+1}$, for which

$$\int |h(c)|^2 \exp\{-|c|^2/2\}\,dc\,dc^* < \infty. \quad (2.1.1.51)$$

Specifically, one-dimensional operators of density $S = |\varphi)(\varphi|$ have kernels

$$k(c', c^*) = (c \mid \varphi)(\varphi \mid c'). \quad (2.1.1.52)$$

2.1.2 Mathematical Models of Sound and Visual Analyzers

In this section we will consecutively introduce and describe mathematical models of an ideal filter, a quasifilter, a disjoint selector, and a quasiselector that make it possible to move to arbitrary representations necessary for solution of the problems of best wave-pattern recognition based on measurements of pattern intensities in a single representation. We will also discuss a theory, based on the work done by Halmos [2.46] and Neumark [2.47] for designing ideal filters and selectors and their realization via indirect measurements, an idea that originated in quantum theory [2.3].

2.1.2.1 Ideal Filters

The simplest measurement of a signal is the determination of the intensity of the oscillations in the signal in a given mode described by a vector ψ normalized to unity, $\|\psi\| = 1$, belonging to the Hilbert space $\mathscr{H}$ of amplitudes φ admissible at the "in" terminals of the receiver and having a finite intensity $I(\varphi) = (\varphi \mid \varphi) < \infty$. The oscillation amplitude in mode ψ is determined by the projection $(\psi \mid \varphi)$ of the received amplitude φ on direction ψ, while the intensity is calculated according to the formula

$$E_\psi(\varphi) = (\varphi \mid \psi)(\psi \mid \varphi) = |(\varphi \mid \psi)|^2, \quad (2.1.2.1)$$

similar to the transition amplitude of a quantum mechanical system from state φ to state ψ. The intensity given by formula (2.1.2.1) is a positive quantity, just as probability is; however, it can assume values greater than unity (but not greater than the total intensity $I(\varphi)$). The appropriate measuring device acts as an ideal filter if it receives a signal φ completely, provided that ψ and φ are collinear,

and does not receive φ if φ and ψ are orthogonal. A mixed signal described by a density operator S excites in mode ψ oscillations of intensity

$$\varepsilon_\psi(S) = \langle S, E_\psi \rangle = (\psi \mid S\psi). \tag{2.1.2.2}$$

More general analyzers carry out the measurement of the intensity

$$E(\varphi) = (\varphi \mid E\varphi) = \| E\varphi \|^2 \tag{2.1.2.3}$$

of the projection $E\varphi$ of the received signal on an arbitrary subspace of $\mathscr{H}$ described by an orthoprojector $E = E^*E$. For example, an audiofrequency filter with a pass band Δ is determined by an orthoprojector $E = I(\Delta)$ on the subspace of amplitudes $\psi(f)$ with support $\Delta = \Phi$ acting as the operator of multiplication by the indicator $1(f, \Delta)$ of set Δ. In a similar manner one can define spatial optical filters that cut out a visual field in a certain region of aperture Δ.

The reader will recall that an orthoprojector is any linear operator E: $\mathscr{H} \to \mathscr{H}$ satisfying the condition $E^* = E = E^2$, and to each Hilbert space $\mathscr{E} \subseteq \mathscr{H}$ there corresponds a unique orthoprojector for which $\mathscr{E} = E\mathscr{H}$. Bearing in mind this one-to-one relation, we will describe such analyzers by orthoprojectors and call them ideal filters, that is, as noted earlier, filters that pass a signal without distortion while measuring its intensity if $\varphi \in E\mathscr{H}$ and that do not pass a signal if it is orthogonal to space $E\mathscr{H}$. The set of all such filters is partially ordered with a smallest element and a greatest element for which we take the null operator 0 and the identity operator I; specifically, filter A is stronger than filter B, or $A \leqslant B$, if orthoprojector B is greater than A, or $AB = A$. Maximal filters, with the exception of the zero filter, are described by one-dimensional orthoprojectors $E_\chi = \mid \chi)(\chi \mid$ acting according to the formula

$$\mid \chi)(\chi \mid \varphi = \chi(\chi \mid \varphi) = (\chi \mid \varphi)\chi \tag{2.1.2.4}$$

and defining the normalized vector χ to within a phase factor $e^{j\theta}$. For every filter A there is a unique complementary filter $A^\perp$ such that $A + A^\perp = I$, with $A^{\perp\perp} = A$, and $A^\perp \geqslant B^\perp$ if $B \geqslant A$. The orthogonal complement $A \to A^\perp$ possesses all the properties of logical negation: if A "passes" φ: $A\varphi = \varphi$, then $A^\perp$ does not: $A^\perp\varphi = 0$, and vice versa, with $A \wedge A^\perp = 0$ and $A \vee A^\perp = I$ with respect to the operations of conjunction $\wedge$ and disjunction $\vee$, for which we take the upper and lower bounds

$$A \wedge B = \sup\{A, B\}, A \wedge B = \inf\{A, B\}, \tag{2.1.2.5}$$

and with the duality formula being valid, or $(A \vee B)^\perp = A^\perp \wedge B$.

Generally, filters A and B are said to be disjoint if $A \perp B$, that is, $A\varphi = \varphi$ implies $B\varphi = 0$ $(B\varphi = \varphi \Rightarrow A\varphi = 0)$, or, which is

equivalent, if $A^{\perp} \geqslant B$ ($B^{\perp} \geqslant A$); they are said to be incompatible if $A \wedge B = 0$, that is, if there is not a single signal that can pass completely through filter A and filter B in the sense that $A\varphi = \varphi$ and $B\varphi = \varphi$.

It can easily be shown that disjoint filters are incompatible, but not the other way round. For this reason the logic of filters is nondistributive, similar to the logic of quantum theory, which is also nondistributive. It satisfies a weaker condition of orthomodularity

$$(A \vee B^{\perp}) \wedge C = A \vee (B^{\perp} \wedge C) \text{ if } A \leqslant B \leqslant C. \tag{2.1.2.6}$$

It is in this respect that the logic of filters differs from Boolean logic, where incompatibility and disjointness mean the same. The intensity of a mixed signal S measured by an ideal filter E can be calculated via the formula

$$\varepsilon(S) = \langle S, E \rangle = \mathrm{Tr}\,(ES). \tag{2.1.2.7}$$

2.1.2.2 Disjoint Selectors

Complicated analyzers measure the intensities $E_i(\varphi) = |(\chi_i \mid \varphi)|^2$ of the received field φ simultaneously in several standard modes $\chi_i \in \mathscr{H}$, $i = 1, \ldots, m$, which, if the normalization conditions $\|\chi_i\| = 1$ for all i's, are necessarily orthogonal in view of the condition $\sum_{i=1}^{m} E_i(\varphi) \leqslant I(\varphi)$. Otherwise, the total received intensity $\sum_{i=1}^{m} E_i(\varphi)$ could be greater than the total intensity $I(\varphi) = \|\varphi\|^2$ of the received signal φ. Such analyzers act as disjoint selectors, or ideal selective filters that split the received signal φ into orthogonal components $\varphi_i = (\chi_i \mid \varphi)\,\chi_i$, $i = 1, \ldots, m$. Signal φ is received completely by a selective filter if $E\varphi = \varphi$, where $E = \sum_{i=1}^{m} E_i$ is the appropriate nonselective filter defined by the one-dimensional orthoprojectors E_i on the subspaces generated by the standard modes χ_i.

More general selective filters are specified by arbitrary sets (or families) $\{E_i \mid i = 1, \ldots, m\}$ of projectors E_i: $\mathscr{H} \to \mathscr{H}$ that satisfy the condition of pairwise orthogonality $E_iE_k = 0$ for $i \neq k$. For example, a disjoint selector that measures the intensity of a signal in each region Δ_i of a Borel partition $\Omega = \sum_i \Delta_i$: $\Delta_i \subseteq \Omega$ is described by an orthogonal set of $E_i = I(\Delta_i)$ of indicators $I(\Delta_i) = \{1(q, \Delta_i)\}$. Note that such a set $\{E_i\}$ may have an infinite number of members if space $\mathscr{H}$ is not finite-dimensional; in this case the re-

ceived intensity is determined for each φ by an absolutely convergent series $\sum_{i=1}^{\infty} E_i(\varphi) \leqslant I(\varphi)$, where

$$E_i(\varphi) = (\varphi \mid E_i\varphi) = \| E_i\varphi \|^2. \tag{2.1.2.8}$$

A selective measurement is said to be complete if the inequality $E(\varphi) \leqslant I(\varphi)$ is transformed into an equality for every $\varphi \in \mathscr{H}$, that is, if $\sum_{i=1}^{\infty} E_i = I$, in a strong operator topology (I is the identity operator in $\mathscr{H}$) and is said to be maximal if all the E_i are one-dimensional.

Complete filters are usually related to self-adjoint operators with a nondegenerate discrete spectrum $\{x_i\}$ through the spectral decomposition (or expansion)

$$A = \sum_{i=1}^{\infty} x_i E_i, \tag{2.1.2.9}$$

with each set $\{E_i\}$ being assigned a numbering self-adjoint operator $N = \sum_i iE_i$.

In addition to discrete filters there is another important class of filters, known as continuous filters, which are related to normal operators with a continuous spectrum $X \subseteq \mathbb{C}^1$. In accordance with von Neumann's theorem, to each such operator there is uniquely assigned a projector-valued measure E on X that specifies the orthogonal expansion (or decomposition) of unity $I = \int E(x)$, so that

$$A = \int xE(dx),\ \mathscr{D}(A) = \left\{\chi \in \mathscr{H}\colon \int |x|^2 E(\chi, dx) < \infty\right\}. \tag{2.1.2.10}$$

Here a family $\{A_j \mid j = 1, \ldots, n\}$ of pairwise commutative normal operators $A_j\colon \mathscr{H} \to \mathscr{H}$ has corresponding to it a selective filter described by a projector-valued measure $E = \bigotimes_{j=1}^{n} E_j$ on $X \subseteq \mathbb{C}^m$ that defines a spectral representation $A = \int xE(dx)$ for the vector operator $A = (A_j)$. It is with these vector selective filters that the measurement of the intensity distribution (2.1.1.1) in the coordinate region is carried out. Such a distribution is described by the orthogonal decomposition of unity $I = \int I(dx)$ for the coordinate vector operator $Q = (Q_k,\ k = 0, 1, \ldots, d)$; the coordinates in the coordinate (or position) representation are given by the respective oper-

ator of multiplication by $q = (q_k)$, so that

$$I(\Delta)\varphi(q) = 1(q, \Delta)\varphi(q), \tag{2.1.2.11}$$

where $1(\Delta)$ is the indicator of the Borel subset $\Delta \subseteq \mathbb{R}^{d+1}$: $1(\Delta, q) = 1$ for $q \in \Delta$ and $1(q, \Delta) = 0$ for $q \notin \Delta$. The result of such a measurement is the continuous measure $I(\varphi, dx)$ with a density

$$\iota(\varphi, x) = I(\varphi, dx)/dx = |\varphi(x)|^2. \tag{2.1.2.12}$$

Note that the self-adjoint position operator

$$Q = \int qI(dq), \quad \mathscr{D}(Q) = \left\{\chi \in \mathscr{H} : \int q^2 |\chi(q)|^2 dq < \infty\right\} \tag{2.1.2.13}$$

has a domain of definition $\mathscr{D}(Q)$ coinciding with $\mathscr{H} = \mathscr{L}^2(Q)$ only in the case of a bounded region Ω. Generally speaking, operator Q is only a densely definite operator, such as the frequency operator $F = \int_0^\infty fI(df)$ in the case of a semi-infinite band $\Phi = [0, \infty[$ of the spectrum.

In general, let X be an arbitrary set and $\mathscr{B}(X)$ the Borel algebra of its subsets. Every measure E: $\Delta \in \mathscr{B} \mapsto E(\Delta)$ with values in the orthoprojectors of the Hilbert space $\mathscr{H}$ is said to be a disjoint selector, it is called a complete selector if $E(X) = I$. Disjoint selectors measure the intensity distribution in the received signal φ on X according to the formula

$$E(\varphi, \Delta) = (\varphi \mid E(\Delta)\varphi) = \| E(\Delta)\varphi \|^2, \tag{2.1.2.14}$$

and define for each φ a positive measure on X of finite mass $E(\varphi, X) \leqslant I(\varphi)$, coinciding with $I(\varphi) = \|\varphi\|^2$ in the case of a complete selector.

We will say that selector E' on X' majorizes selector E on X ($E' \succcurlyeq E$) if there exists a measurable mapping f: $X' \to X$ with respect to which

$$E(\Delta) = E'(f^{-1}(\Delta)) \quad \text{for every } \Delta \in \mathscr{B}(X) \tag{2.1.2.15}$$

where $f^{-1}(\Delta) = \{x' \in X' \mid f(x') \in \Delta\}$ is the inverse image of set $\Delta \subseteq X$, and the selectors E' and E are equivalent, $E' \simeq E$, if $E' \preccurlyeq E$, too.

2.1.2.3 Successive Filters and Quasifilters

The common practice in processing sound and visual patterns is to use analyzers that act not in the initial Hilbert space $\mathscr{H}$ generated by the amplitudes φ on the "in" terminals but in an extension of this space. For example, temporal measurements of sound

signals $\varphi(f)$ with a restricted frequency band Φ are reduced to determining the intensity of these signals in this or that temporal interval $\Delta \in \mathbb{R}^1_+$ via the orthoprojector $\tilde{I}(\Delta)$ of multiplication of the signal in the temporal representation by the indicator function $1(t, \Delta)$:

$$\tilde{I}(\varphi, \Delta) = \int 1(t, \Delta) \, |\tilde{\varphi}(t)|^2 dt = \|\tilde{I}(\Delta)\varphi\|^2 \qquad (2.1.2.16)$$

where $\tilde{I}(\Delta)$ acts in the space of signals of unlimited bandwidth, $\mathcal{L}^2(\mathbb{R})$. In a similar manner space $\mathcal{H}$ is extended to $\mathcal{L}^2(\mathbb{R}^d)$ in the wave processing of optical fields observed on a limited aperture $S \subset \mathbb{R}^d$, which results in determining the intensities of the fields in this or another momentum interval $\Delta \in \mathbb{R}^d$.

In general, such an extension is described by an isometric embedding F: $\mathcal{H} \to \mathcal{H}'$ of Hilbert space $\mathcal{H}$ into another space $\mathcal{H}'$, for which for examples considered here we can take the Hilbert space $\mathcal{H}' = \mathcal{L}^2(\mathbb{R}^{d+1})$ into which the space $\mathcal{H} = \mathcal{L}^2(\Omega)$ is isometrically embedded via the Fourier transform F: $\varphi \mapsto \tilde{\varphi}$.[2] An ideal filter described in $\mathcal{H}'$ by the orthoprojector E measures the intensity of amplitude $\varphi \in \mathcal{H}$ defined by the Hermitian form

$$D(\varphi) = \|EF\varphi\|^2 = (F\varphi \mid EF\varphi) = (\varphi \mid D\varphi), \quad (2.1.2.17)$$

where $D = F^*EF$ is a positive contraction operator in $\mathcal{H}$. Formula (2.1.2.17) shows that this intensity can be considered the result of successive action of two ideal filters, F and E, with $\mathcal{H}$ being identified with a subspace $F\mathcal{H} \subset \mathcal{H}'$, where filter F is described by the orthoprojector F that cuts subspace $\mathcal{H}$ out of $\mathcal{H}'$. For example, temporal measurement of narrow-band signals is the result of non-commutative action of a frequency filter $F = I(\Delta f)$ and a temporal filter $E = \tilde{I}(\Delta t)$, the result is effectively described by the Hermitian form (2.1.2.17) defined by the operator $D = I(\Delta f)\,\tilde{I}(\Delta t)\,I(\Delta f)$. It is, therefore, advisable to generalize the concept of a filter by describing it in space $\mathcal{H}$ by any operator D: $\mathcal{H} \to \mathcal{H}$ that satisfies the condition

$$I \geqslant D^* = D \geqslant 0, \qquad (2.2.1.18)$$

and calling it a quasifilter if $D \neq D^2$.

The basis for this extension is the Halmos theorem [2.46], according to which every quasifilter described by operator (2.1.2.18) can be considered as a reduction (projection) $D = F^+EF$ on $\mathcal{H}$ of an ideal filter E acting in an extension $\mathcal{H}'$. For the Hilbert space $\mathcal{H}'$ we can always take the doubling $\mathcal{H}' = \mathcal{H} \oplus \mathcal{H} = \mathbb{C}^2 \otimes \mathcal{H}$ of space $\mathcal{H}$

[2] The isometry F may be antilinear rather than linear, as is the case with the involution Fourier transform (2.1.1.5).

with embedding F: $\varphi \mapsto (\varphi, 0)$, selecting the operators

$$E_{11} = D, \; E_{12} = \sqrt{D(1-D)} = E_{21}, \; E_{22} = I - D \quad (2.1.2.19)$$

for the blocks of orthoprojector E. Note that allowance for consecutive action of several noncommutative ideal filters $E_1, \ldots, E_n$ in the initial Hilbert space also leads to the notion of a quasifilter. These ideal filters then measure the intensity

$$D(\varphi) = \| E_n \ldots E_1 \varphi \|^2 = (\varphi \mid D\varphi), \quad (2.1.2.20)$$

$$D = E_1 \ldots E_{n-1} E_n E_{n-1} \ldots E_1, \quad (2.1.2.21)$$

and the result can be considered the effect of linear nonideal filters not necessarily described by Hermitian contraction operators A: $\mathscr{H} \to \mathscr{H}$, $\| A \| \leqslant 1$, with the nonideal filters damping and distorting the amplitudes and with $D = A^*A$ in the formula for the appropriate intensity:

$$D(\varphi) = \| A\varphi \|^2 = (\varphi \mid D\varphi). \quad (2.1.2.22)$$

2.1.2.4 Quasiselectors and Indirect Measurements

In a similar manner we can introduce generalized selectors, which are defined on a Borel space X by a positive operator-valued measure M: $\Delta \in \mathscr{B}(X) \to M(\Delta)$ specifying in the Hilbert space $\mathscr{H}$ a weak decomposition $D = \int M(dx)$ of an operator D satisfying condition (2.1.2.18) in the following sense:

$$D(\varphi) = \int M(\varphi, dx) \text{ for every } \varphi. \quad (2.1.2.23)$$

Here, as usual, $D(\varphi) = (\varphi \mid D\varphi)$ is a Hermitian form defined by the operator of total effect D, while

$$M(\varphi, dx) = (\varphi \mid M(dx)\, \varphi) \quad (2.1.2.24)$$

is the distribution of intensity of X corresponding to amplitude φ and measured by such a selector. Note that the expansion of operator D defined by measure $M(dx)$ may not necessarily be orthogonal even if the operator is a projector, that is, if the total filter is ideal, such nondisjoint selective filters will be called quasiselective filters, or simply quasiselectors. A quasiselector is said to be complete if $M(X) = I$ and maximal if $M \leqslant M' \Rightarrow M' \simeq M$ in the same sense as in (2.1.2.15).

An example of a complete maximal quasiselector for sound signals and optical fields observed in a restricted region Ω of coordinates $\mathbb{R}^{d+1} \ni q = (t, \mathbf{q})$ is the analyzer of the momentum distribution (2.1.1.4), which in the q-representation is defiined by formula (2.1.2.24)

via the operator-valued measure

$$M(dp) = F^* I(dp) F \equiv \tilde{I}(dp), \tag{2.1.2.25}$$

where F is the Fourier transform (2.1.1.5), and $I(D) = \{1(p, D)\}$ is the projector-valued measure on $\mathbb{R}^{d+1}$ described in the space $\mathcal{H}' = \mathcal{L}^2(\mathbb{R}^{d+1})$ of the proper representation of the generalized-momentum operator $p = (t, \mathbf{p})$ by the indicator measure $1(p, dx)$.

Note that the momentum operator defined in $\mathcal{H} = \mathcal{L}^2(\Omega)$ by the nonorthogonal expansion

$$P| = \int p\tilde{I}(dp), \quad \mathcal{D}(P) = \left\{\chi \in \mathcal{H} |: \int p^2 \,|\, \tilde{\chi}(p)\,|^2\, dp < \infty\right\}$$

is always unbounded with a spectrum $\mathbb{R}^{d+1}$ and, for $\Omega \neq \mathbb{R}^{d+1}$, non-self-adjoint, notwithstanding the fact that the form of the total momentum $P(\varphi) = \int \tilde{I}(\varphi, dp)$ is always Hermitian. Nevertheless, this operator always uniquely defines a nonorthogonal expansion $I = \int \tilde{I}(dp)$ via the condition

$$\int p^2 (\chi \mid \tilde{I}(dp)\,\chi) = (P\chi \mid P\chi), \quad \chi \in \mathcal{L}(P),$$

and is a restriction to functions $\varphi(q) = 0$ for $q \notin \Omega$ of the operator $(2\pi j)^{-1}\, \partial/\partial q$ that is self-adjoint in $\mathcal{L}^2(\mathbb{R}^{d+1})$ with a domain of definition $\{\varphi \in \mathcal{L}^2(\mathbb{R}^{d+1})\colon \|\partial^2\varphi/\partial q^2\|^2 < \infty\}$. For instance, time operator $T = \int t\tilde{I}(dt)$ in the space $\mathcal{H} = \mathcal{L}^2(\Phi)$ with a semi-infinite band $\Phi = [0, \infty[$ is a symmetric but not a self-adjoint operator $(2\pi j)^{-1}\,\partial/\partial f$ with a domain of definition

$$\mathcal{D}(T) = \left\{\chi \in \mathcal{L}^2(0, \infty) : \chi(0) = 0, \int_0^\infty |\,\partial\chi(f)/\partial f\,|^2\, df < \infty\right\}.$$

Similarly, the validity of the representation

$$M(dx) = F^* E(dx) F, \quad F^* F = I, \tag{2.1.2.26}$$

for an arbitrary quasiselector M in the form of the projection of the disjoint selector E described in the extended space $\mathcal{H}'$ by an orthogonal projector-valued measure $E(dx)$ is ensured by the Neumark theorem [2.47]. For mixed signals the intensity, as a function of the density operator S, is described by a measure $\mu(S, dx)$ defined by the following linear form:

$$\mu(S, dx) = \langle S, M(dx)\rangle. \tag{2.1.2.27}$$

Quasiselective filters also emerge as a result of reducing the description of indirect measurement of the received signal via the disjoint selection $E_0(dx)$ of the initially uncorrelated reference signal

interacting with the received signal; this reference signal generates a Hilbert space $\mathcal{H}_0$. Specifically, if S_0 is the density operator of the normalized reference signal ($\mathrm{Tr}\, S_0 = 1$) and U is a unitary operator describing in the tensor product $\mathcal{H} \otimes \mathcal{H}_0$ the result of the interaction $S' = U(S \otimes S_0)U^*$ with the received signal S, then the intensity distribution corresponding to such indirect measurement may be effectively calculated via formula (2.1.2.27) as a result of the quasi-measurement

$$\langle S, M(dx) \rangle = \langle S', I \otimes S_0(dx) \rangle = \langle S \otimes S_0, E'(dx) \rangle$$

described by the operator-valued measure

$$M(dx) = \mathrm{Tr}\, \{(I \otimes S_0)\, E'(dx) \mid \mathcal{H}\}, \tag{2.1.2.28}$$

where $E'(dx) = U^* (I \otimes E_0(dx))\, U$, and $\mathrm{Tr}\, \{\cdot \mid \mathcal{H}\}$ is the partial trace in $\mathcal{H} \otimes \mathcal{H}_0$ defined for factorable density operators $S \otimes S_0$ via the formula

$$\mathrm{Tr}\, \{S \otimes S_0 \mid \mathcal{H}\} = S\, \mathrm{Tr}\, S_0.$$

The indirect calculation of the intensity distribution over the frequency (color) $f \in \Phi$ of static monochrome patterns is an example of the above-mentioned type of measurement. It can be carried out as a result of the wave processing of the patterns in which the intensity distribution over the momenta p in such patterns is calculated.

Using the Neumark theorem as a basis, let us give an explicit description of a construction that makes it possible to reduce any quasimeasurement to an indirect measurement. To this end we take for $\mathcal{H}_0$ the space $\mathcal{H}'$ of the Neumark construction and for the reference signal a normalized amplitude $\psi' = F\psi$, $\|\psi\| = 1$, and introduce the linear operator U in $\mathcal{H} \otimes \mathcal{H}'$ (which is defined by the Neumark isometry $F: \mathcal{H} \to \mathcal{H}'$, $F^*F = I$) in the following manner:

$$U: \varphi \otimes \varphi' \mapsto F^*\varphi' \otimes F\varphi + \varphi \otimes (1 - FF^*)\, \varphi' \tag{2.1.2.29}$$

with the generating elements being $\varphi \otimes \varphi'$, $\varphi \in \mathcal{H}$, $\varphi' \in \mathcal{H}'$. It can be directly verified that $U = U^*$ and $U^2 = I$ and, hence, $U^*U = I = UU^*$. Taking for $E_0(dx)$ the Neumark expansion $E(dx)$ in $\mathcal{H}'$, assuming that $S_0 = |\psi)(\psi|$ and allowing for (2.1.2.26), we arrive at an indirect measurement whose reduction (2.1.2.28) yields the initial measure $M(dx)$:

$$\mathrm{Tr}\, \{(I \otimes |\psi')(\psi'|)\, E'(dx)\} = (\psi \mid \psi)\, F^*E(dx)\, F$$
$$= M(dx).$$

2.1.2.5 Canonical Operators and Measurements

Bearing in mind the invariance of the domains of definitions of operators Q and P with respect to the self-adjoint operators of multiplication by q and differentiation with respect to q, or $(2\pi j)^{-1}\,\partial/\partial q$, which on $\mathscr{H} = \mathscr{L}^2\,(\Omega)$ coincide with Q and P, respectively, in what follows we will take for Q and P in $\mathscr{L}^2\,(\mathbb{R}^{d+1})$ their extensions, while always assuming that the region where these operators act is $\Omega \subset \mathbb{R}^{d+1}$. Such operators are known as canonical and satisfy the commutation relations (2.1.1.9) in the common domain $\mathscr{D}\,(P) \cap \mathscr{D}\,(Q)$.

Let us now discuss simultaneous measurement of the coordinate (or position) and momentum distributions. In view of the noncommutativity of Q and P, there can be no joint orthogonal decomposition of unity for these operators; there can even be no joint nonorthogonal decomposition of $M\,(dx\,dy)$ for which the following would be true:

$$I\,(dq) = \int M\,(dq\,dy),\quad \tilde{I}\,(dp) = \int M\,(dx\,dp). \tag{2.1.2.30}$$

Otherwise, in view of the Neumark theorem, there would be commutative self-adjoint operators in $\mathscr{H}' \supset \mathscr{H}$ coinciding on $\mathscr{H}$ with the noncommutative operators Q and P, which is impossible.

Another interesting question is the relation to these operators of the measurements of the canonical distributions (2.1.1.39). Such canonical measurements are described, obviously, by continuous with respect to $dz = dx\,dy$ nonorthogonal measures $K\,(dz) = k\,(z)\,dz$ with projector-valued densities

$$k\,(z) = |\,\psi_z)\,(\psi_z\,| = |\,c)\,(c\,| \equiv k\,(c,\ c^*), \tag{2.1.2.31}$$

which are defined by canonical amplitudes (2.1.1.25) at $\zeta = z$ and a certain ω or, in complex variables, by (2.1.1.34) at $\alpha = c$. The respective quasiselective filters, which are parameterized by symmetric ω matrices with a nonsingular real part $\omega + \omega^*$ and now will be called canonical filters, are, obviously, maximal and, because of (2.1.1.42), complete:

$$\int |\,\psi_z)\,(\psi_z\,|\,dz = I = \int\int |\,c)\,(c\,|\,dc\,dc^*, \tag{2.1.2.32}$$

where $dc\,dc^* = dx\,dy$.

By fixing ω and directly integrating we find that the quasimeasurement of an intensity distribution in $x \in \mathbb{R}^{d+1}$ is described by a continuous measure $M\,(dx) = \int K\,(dx\,dy) = m\,(x)\,dx$ diagonal in the q-representation,

$$m\,(x) = \int k\,(x,\ y)\,dy = |\,\varepsilon\,|^{-1} \int e^{-\pi\,|\,(x-q)\varepsilon^{-1}\,|^2}\,I\,(dq). \tag{2.1.2.33}$$

with a Gaussian density and $\upsilon^+\upsilon = 2\pi(\omega + \omega^*)^{-1}$, while the quasi-measurement of an intensity distribution in $y \in \mathbb{R}^{d+1}$ is described by an operator measure $\widetilde{M}(dy) = \int K(dx\,dy) = \widetilde{m}(y)\,dy$, where

$$\widetilde{m}(y) = \int k(x, y)\,dx = |\widetilde{\upsilon}|^{-1} \int e^{-\pi|(y-p)\widetilde{\upsilon}^{-1}|^2} \widetilde{I}(dp), \tag{2.1.2.34}$$

with a Gaussian density and $\widetilde{\upsilon}^+\widetilde{\upsilon} = 2\pi(\widetilde{\omega} + \widetilde{\omega}^*)^{-1}$, where $\widetilde{\omega}/2\pi = 2\pi/\omega^*$. The operator measures M and $\widetilde{M}$ on $\mathbb{R}^{d+1}$, which define nonorthogonal expansions of operators Q and P, that is,

$$Q = \int xm(x)\,dx \text{ and } P = \int y\widetilde{m}(y)\,dy, \tag{2.1.2.35}$$

describe, in contrast to the spectral measures I and $\widetilde{I}$, inaccurate measurements of position and momentum distributions, which are obtained by smoothing out (2.1.1.40) and (2.1.1.41) with Gaussian weighting functions m and $\widetilde{m}$. Nevertheless, the canonical operator measure K that generates the two spectral measures possesses certain spectral properties with respect to the complex-valued combinations of the two respective operators:

$$A = (Q\omega + 2\pi jP)\,\upsilon^+/\sqrt{2\pi}. \tag{2.1.2.36}$$

Namely, applying A directly to the canonical amplitudes (2.1.1.34), we can easily verify that it is well-defined on these amplitudes:

$$A\,|\,\alpha) = \alpha\,|\,\alpha),\ \alpha \in \mathbb{C}^{d+1}, \tag{2.1.2.37}$$

which, therefore, form a proper base for A in $\mathcal{H} = \mathcal{L}^2(\mathbb{R}^{d+1})$.

Hermitian conjugate operators $C = A^*$ are diagonal in the Bargmann representation,

$$(\hat{c}h)(c) = e^{|c|^2/2}(C\chi\,|\,c) = e^{|c|^2/2}c(\chi\,|\,c) = ch(c),$$

with a domain of definition $\mathcal{D}(\hat{c}) = \{\chi \in \mathcal{H} : \|\hat{c}h\| < \infty\}$, where

$$\|\hat{c}h\|^2 = \int |c|^2\,|h(c)|^2 e^{-|c|^2} dc\,dc^*, \tag{2.1.2.38}$$

on which domain there is also defined the operator A by differentiation $(\hat{a}h)(c) = \partial h(c)/\partial c$. Thus, in the initial representation we obtain the spectral nonorthogonal decomposition

$$A^* = \int c^* K(dz),\ \mathcal{D}(A^*)$$
$$= \left\{\chi \in \mathcal{H} : \int |c|^2\,|(\chi\,|\,c)^2 dc\,dc^* < \infty\right\}.$$

Now let us describe a simple realization of a canonical measurement by an indirect measurement defined in the tensor product $\mathcal{H} \otimes \mathcal{H}_0$, where $\mathcal{H}_0$ is a copy of $\mathcal{H}$. To this end we take the commutative self-adjoint operators

$$X = Q \otimes I_0 + I \otimes Q_0,\ Y = P \otimes I_0 - I \otimes P_0, \qquad (2.1.2.39)$$

where $Q_0 = q_0$, $P_0 = (2\pi j)^{-1}\, \partial/\partial q_0$, and I_0 is the identity operator in $\mathcal{H}_0 = \mathcal{L}^2(\mathbb{R}^{d+1})$. Suppose that $E(dz)$ is the orthogonal spectral measure of the set $Z = (X, Y)$ and that

$$\psi_0(q_0) = |(\omega + \omega^*/(2\pi)|^{1/4} \exp\left\{-\frac{1}{2} q_0 \omega q_0^{\mathrm{T}}\right\} = |0)_0 \qquad (2.1.2.40)$$

is the basic canonical amplitude in $\mathcal{H}_0$. We take an arbitrary amplitude $\chi \in \mathcal{H}$ and the corresponding tensor product $(\chi \otimes \psi_0^*)(q, q_0) = \chi(q)\psi_0^*(q_0)$ in $\mathcal{H} \oplus \mathcal{H}_0$ and define the characteristic function of the corresponding distribution thus:

$$\Upsilon(u, u^*) = \int e^{j(u_* c^{\mathrm{T}} + c_* u^{\mathrm{T}})} (\chi \otimes \psi_0^* \mid E(dz)\, \chi \otimes \psi_0^*), \qquad (2.1.2.41)$$

$$z = (x, y) \in \mathbb{R}^{2(d+1)}$$

where as usual, $c = (x\omega + 2\pi j y)\, \upsilon^+ \sqrt{2\pi}$ and $u, u^* \in \mathbb{C}^{2(d+1)}$. We write this function in terms of normal operators

$$B = \int cE(dz) = (X\omega + 2\pi j Y)\, \upsilon^+/\sqrt{2\pi} = A \otimes I_0 + I \otimes C_0, \qquad (2.1.2.42)$$

with A the operators (2.1.2.36) and $C_0 = (Q_0\omega - 2\pi j P_0)\, \upsilon^+/\sqrt{2\pi}$, in the form

$$\Upsilon(u, u^*) = (e^{jB_* u^{\mathrm{T}}} \chi \otimes \psi_0^* \mid e^{jB_* u^{\mathrm{T}}} \chi \otimes \psi_0^*) = (e^{jA_* u^{\mathrm{T}}} \chi \mid e^{jA_* u^{\mathrm{T}}} \chi),$$

where we have allowed for the property $C_0 \psi_0^* = 0$ for the basic amplitude (2.1.2.40). Employing now the completeness property (2.1.2.32) of canonical amplitudes, we obtain

$$\begin{aligned}\Upsilon(u, u^*) &= \int (e^{jA_* u'} \chi \mid c)(c \mid e^{jA_* u'} \chi)\, dc\, dc^* \\ &= \int e^{ju_* c' + jc_* u'} |(\chi \mid c)|^2\, dc\, dc^* = \widetilde{k}(u, u^*).\end{aligned} \qquad (2.1.2.43)$$

Thus, the characteristic function (2.1.2.41) of the indirect measurement of intensity of amplitude $\chi \in \mathcal{H}$ coincides for the ground state

ψ_0^*, when calculated in the z-representation of operators (2.1.2.39), with the characteristic function $\tilde{k}$ of the canonical distribution $k(c, c^*) = |(\chi \mid c)|^2$ for this amplitude.

2.2 Optimal Detection and Discrimination of Acoustic Signals and Optical Fields

In this chapter we develop the wave theory of hypothesis testing for solving problems of optimal recognition of sound and visual patterns. We formulate the necessary and sufficient conditions for the optimality of two-alternative and multialternative detection of wave patterns according to the maximum criterion for the measured intensity of acoustic signals and optical fields. We consider problems involving the discrimination of a wave pattern against an acoustic or optical background, problems involving the discrimination of pure nonorthogonal signals and fields, and problems involving the recognition of mixed patterns described by noncommutative density operators. Complete solution of the last type of problem is then obtained for the case of mixing two pure patterns. The discussed results of solution of the corresponding extremal problems follow from the methods of linear programming in Banach partially ordered operator spaces [2.48]. These results generalize the corresponding results of the quantum detection and estimation theory, which have been obtained for the two-alternative case by Helstrom [2.11] and for the multialternative case by Belavkin [2.4, 2.5]. The necessary and sufficient optimality conditions for the quantum theory of hypothesis testing have been discussed by Kennedy [2.9], Yuen and Lax [2.15], Kholevo [2.24], Belavkin [2.4], and Belavkin and Vancjan [2.27].

2.2.1 Optimal Detection of Sound and Visual Patterns

In this section we will discuss the problem of detecting wave patterns that are in a partially coherent superposition with an acoustic or optical background. The problem is complicated by the presence of interference. We start by considering the superposition principle for generalized mixed amplitudes. We then formulate the necessary and sufficient conditions for the optimality of detection and give solutions to a number of problems considered in the quantum case in the review [2.30].

2.2.1.1 The Superposition Principle

The problem of detecting a sound or visual pattern described by a wave amplitude $\varphi(q)$ taken from the Hilbert space $\mathcal{H} = \mathcal{L}^2(\Omega)$ can be solved in a trivial manner by measuring the total intensity $I(\varphi) = \|\varphi\|^2$ only in the absence of an acoustic

or optical background consisting of other signals and fields in the frequency-spatial region Ω considered. If in the region of measurement there is another signal or field described by amplitude $\varphi_0 \in \mathscr{H}$ the question of whether a wave pattern φ is present cannot generally be unambiguously solved by simply measuring the total intensity of the resulting amplitude ψ, which may be higher or lower than the background intensity. Such a phenomenon is called interference and is the result of the wave superposition principle $\psi = \varphi + \varphi_0$, according to which the complex-valued amplitudes of the coherent signals φ and φ_0 rather than the intensities of these signals, are added. The intensity of the resulting signal has the form

$$\| \psi \|^2 = \| \varphi \|^2 + 2 \operatorname{Re} (\varphi \mid \varphi_0) + \| \varphi_0 \|^2. \qquad (2.2.1.1)$$

To describe the result of the superposition of a mixed pattern and a partially coherent background caused, say, by thermal fluctuations that have an infinite total intensity of the acoustic or optical field, we can employ the correlation theory by considering generalized random amplitudes within the second-order statistical theory.

Partially coherent signals and fields determined in a similar manner in the framework of the classical or the quantum theory are commonly described by bounded operators F from $\mathscr{H}$ to another Hilbert space $\mathscr{K}$; for ordinary nonrandom amplitudes $\psi \in \mathscr{H}$ these operators are usually represented by the functionals $F_\psi \chi = (\psi \mid \chi)$, denoted by $F_\psi = (\psi \mid$ and acting from $\mathscr{H}$ to $\mathscr{K} = \mathbb{C}$. The mean intensity of random oscillations excited in mode $\chi \in \mathscr{H}$, $\| \chi \| = 1$, is determined by a Hermitian form in F:

$$E_\chi (F) = (F\chi \mid F\chi) = (\chi \mid F^+F\chi), \qquad (2.2.1.2)$$

and is calculated for common mixed signals via formula (2.1.2.2) with the aid of a (generally infinite trace) density operator $P = F^+F$ of the intensity $\iota(P) \in [0, \infty]$, where F^+ is the Hermitian conjugate operator $\mathscr{K} \to \mathscr{H}$ acting for $F = (\psi \mid$ as an operator of multiplication $c \mapsto \psi c$ from $K = \mathbb{C}$ to $\mathscr{H}$. The intensity (2.2.1.2) is a measurable quantity bounded by the norm of the positive operator P and equal, via the duality theorem, to

$$\varepsilon_\chi (P) = (\chi \mid P\chi) \leqslant \| P \| = \inf \{\varepsilon \mid \varepsilon I \geqslant P\}, \qquad (2.2.1.3)$$

which for the case of white noise $P = \varepsilon I$, described by the isometric operator $T = F/\sqrt{\varepsilon}$, $T^+T = I$, determines the local intensity $\varepsilon = \varepsilon_\chi (P)$, the same for all modes $\chi \in \mathscr{H}$. Note that every partially coherent signal F can be considered as the result of action of a contraction filter $D = P/\| P \|$ on white noise of local intensity $\varepsilon = \| P \|$ if we employ the polar decomposition $F = TP^{1/2}$, which determines uniquely the isometry operator T on the range of values $F^+\mathscr{K}$ of operator F^+.

For generalized signals with infinite trace density operators P it proves expedient, however, to consider only such quasimeasurements for which the operators D of the total effect lead to finite intensities:

$$D(F) \equiv \mathrm{Tr}\,(FDF)^{+} = \mathrm{Tr}\,(PD). \qquad (2.2.1.4)$$

In addition to one-dimensional projectors $D = E_\chi = |\chi)(\chi|$, for which the intensity (2.2.1.2) is determined by the bounded form (2.2.1.3), we can always consider finite-dimensional operators $D = \sum_i |\chi_i)(\chi_i|$ as well as any trace class operator $0 \leqslant D < I$, since $\mathrm{Tr}\,(SD) \leqslant \varepsilon\,\mathrm{Tr}\,D$ if $S < \varepsilon I$.

Extending the superposition principle to generalized amplitudes F, F_0: $\mathscr{H} \to \mathscr{K}$, we find that the result $G = F + F_0$ of addition of the generalized signal F and the background F_0 is described by a density operator $R = G^+G$, that is the sum of operators $P = F^+F$ and $P_0 = F_0^+F_0$ only if $\mathrm{Re}\,F^+F_0 = 0$. The latter condition, which defines the incoherence relation between F and F_0, cannot be met for nonrandom amplitudes $F = (\varphi|$ and $F_0 = (\varphi_0|$ since $F^+F_0 = |\varphi)(\varphi_0| \neq 0$ even in the event of orthogonality $(\varphi|\varphi_0) = 0$ if $\varphi \neq 0$ or $\varphi_0 \neq 0$, although the total intensity (2.2.1.1) is equal to the sum $\|\varphi\|^2 + \|\varphi_0\|^2$.

Generally, the resulting density operator R can be represented in the form

$$R = P + P^{1/2}\Gamma P_0^{1/2} + P_0^{1/2}\Gamma^+P^{1/2} + P_0, \qquad (2.2.1.5)$$

where $\Gamma = T^+T_0$ is the operator of mutual coherence of signal $F = TP^{1/2}$ and noise $F_0 = T_0P^{1/2}$ determined by the partial isometries T: $F^+\mathscr{K} \to \mathscr{H}$ and T_0: $F_0^+\mathscr{K} \to \mathscr{H}$.

Note that Γ is a contracting operator: $\|\Gamma\| \leqslant \|T^+\|\,\|T_0\| = 1$, and a partially isometric operator if $F_0\mathscr{H} \subseteq F\mathscr{H}$. The latter condition determines the coherence relation between the generalized amplitude F_0 and amplitude F, which is always met for nonrandom amplitudes $F = (\varphi|$ and $F_0 = (\varphi_0|$ for which $F_0\mathscr{H} = \mathbb{C} = F\mathscr{H}$. Representing the partially coherent amplitude F_0 in the form of a sum of the component $H_0 = T\Gamma P_0^{1/2}$ coherent with F and the component $W = F_0 - H_0$ that is incoherent and doing the same with the resulting amplitude G, or $G = H_1 + W$, where $H_1 = F + H_0$, we can isolate from operators P_0 and R a common density operator of the incoherent background $N = W^+W$ by writing the two operators, with allowance made for the fact that $W^+H_i = 0$, in the form $P_0 = S_0 + N$ and $R = S_1 + N$, where $S_i = H_i^+H_i$ are the operators $S_0 = P_0^{1/2}\Gamma^+\Gamma P_0^{1/2}$ and

$$S_1 = P + P^{1/2}US_0^{1/2} + S_0^{1/2}U^+D^{1/2} - S_0, \qquad (2.2.1.6)$$

with U the partially isometric operator of polar expansion, and $\Gamma P_0^{1/2} = US_0^{1/2}$. In contrast to P_0 and R, for a trace class operator P the operators S_i are usually also trace class operators of rank $r(S_i) \leqslant r(P)$ and one-dimensional operators if $r(P) = 1$.

Infinite trace operators P may also be replaced with trace class operators if we consider finite total intensities (2.2.1.4) with respect to a fixed D, assuming that $S = D^{1/2}PD^{1/2}$. The effective operator D is then replaced with the orthoprojector E on the subspace $\mathscr{E} = D\mathscr{H}$ that determines the total intensity $\varepsilon(S) = \mathrm{Tr}\, S$ by taking the trace $\varepsilon(S) = \mathrm{Tr}(ES)$ on $\mathscr{E}$.

2.2.1.2 Classical Detection

The simplest detection problem, that of isolating a pattern described by a kernel operator $P > 0$ from an incoherent mixture $R = P + N$ of this pattern with the background N, is solved by measuring the intensity of one of the possible signals, $R_0 = N$ or $R_1 = R$, by comparing this signal with the background level $\iota(N) = \mathrm{Tr}\, N$. To this end it has proved sufficient to limit oneself to measuring the total degree of contrast $\iota(C) = \mathrm{Tr}\, C$ of the received signal by calculating the trace $\langle C, E\rangle = \mathrm{Tr}(CE)$ of the appropriate operator $C_i = R_i - N$, $i = 0, 1$, on any subspace $\mathscr{E} = E\mathscr{H}$, $CE = C$, with the trace assuming finite values $\langle C_0, E\rangle = 0$ in the absence of a pattern, $i = 0$, and $\langle C_1, E\rangle = \mathrm{Tr}\, P$, $i = 1$, in the presence of a pattern even for an infinitely high level of the background $\iota(N) = \infty$.

In the case of a partially coherent superposition R of pattern P and background P_0, the difference $C = R - P_0$ may be a nonpositive trace class operator with a zero or even negative trace, with the result that the detection criterion, which is based on the condition that the total degree of contrast $\iota(C)$ is positive, may lead to incorrect results. Even if $\iota(C)$ is positive, which is the case when the superposition $\psi = \varphi_0 + \varphi$ of orthogonal amplitudes, $(\varphi \mid \varphi_0) = 0$, is coherent, that is, $\iota(C) = \|\psi\|^2 - \|\varphi_0\|^2 = \|\varphi\|^2$, we can considerably increase the degree of contrast of amplitudes φ_0 and ψ if we sum, say, the coordinate distribution of the degree of contrast,

$$c(x) = |\psi(x)|^2 - |\varphi_0(x)|^2 = |\varphi(x)|^2 + 2\,\mathrm{Re}\,\varphi^*(x)\,\varphi_0(x), \tag{2.2.1.7}$$

not over the entire region Ω but only that part of the region where $c(x)$ is positive. As a result we arrive at the following classical problem of optimal detection of a pattern in a coordinate (frequency-spatial) region Ω: we must find a measurable subregion $\Delta^0 \subseteq \Omega$ in which the upper bound

$$\varkappa_1^0(C) = \sup_{\Delta \subseteq \Omega} \langle C, I(\Delta)\rangle = \int_{\Delta^0} c(x)\,dx \tag{2.2.1.8}$$

of the integral of the contrast function $c(x) = C(x, x)$ is attained. This function is determined by the diagonal values of the kernel $C(x', x)$, which is the difference between the generalized matrix elements $R(x', x)$ and $N(x', x)$ of operators R and N in the coordinate representation.

It is sufficient to consider the supremum (2.2.1.8) in the class of measurable subsets $\Delta \subseteq \Omega$ of the coordinate region $\Omega = \{x \in \mathbb{R}^{d+1} \mid c(x) \neq 0\}$, the support of the integrable function $c(x)$, in which the supremum is attained only on the set

$$\Delta^0 = \{x \in \Omega \mid c(x) > 0\} \equiv \Omega_+. \qquad (2.2.1.9)$$

Its value, $\varkappa_I^0(C) = \int_{\Omega_+} c(x)\,dx$, coincides, obviously, with the integral over Ω of the positive part

$$c_+(x) = \max\{0, c(x)\} = \frac{1}{2}(c(x) + |c(x)|), \qquad (2.2.1.10)$$

where the functions c determine the solution to the duality problem

$$\langle c\rangle_+ = \inf_{b \geqslant 0}\left\{\int_\Omega b(x)\,dx \mid b \geqslant c\right\} = \int_\Omega c_+(x)\,dx. \qquad (2.2.1.11)$$

The lower bound (2.2.1.11) over all positive integrable functions $b(x) \in \mathscr{L}_+^1(\Omega)$, majorizing almost everywhere the function c, is attained at $b^0 = 0 \vee c = c_+$ and determines on the space of integrable functions c a positive gauge $\langle c\rangle_+$, which is zero only when $c \leqslant 0$. The set (2.2.1.9) specifies the optimal band of the frequency-spatial filter in which the best quality of detection, (2.2.1.8), is achieved.

Reasoning along similar lines, we can solve the problem of optimal detection in the momentum (or temporal-wave) space $X = \mathbb{R}^{d+1}$,

$$\varkappa_I^0(C) = \sup_{\Delta \subseteq X}\langle C, \tilde{I}(\Delta)\rangle = \int_{\Delta^0} \tilde{c}(x)\,dx, \qquad (2.2.1.12)$$

where $\tilde{c}(x) = \tilde{C}(x, x)$ are the diagonal elements of the difference $\tilde{R}(x', x) - \tilde{N}(x', x)$ of the operators R and N in the momentum representation; in coherent superposition these diagonal elements are

$$\tilde{c}(x) = |\tilde{\psi}(x)^2 - |\tilde{\varphi}_0(x)|^2 = |\varphi(x)|^2 + 2\,\mathrm{Re}\,\tilde{\varphi}^*(x)\,\tilde{\varphi}_0(x). \qquad (2.2.1.13)$$

The quality of such detection, $\varkappa_{\tilde{I}}^0(C) = \langle\tilde{C}\rangle_+$, based on a momentum quasimeasurement may differ considerably from (2.2.1.11). For example, the canonical amplitudes (2.1.1.25) $\varphi_0 = \psi_{00}$ and $\psi = \psi_{0\eta}$, which are similarly localized in the coordinate representation, differ by their momenta, $\eta \neq 0$, and can be thought of as two hypotheses, corresponding to the absence and presence of a complex-

valued amplitude $\varphi = \psi_{0\eta} - \psi_{00}$ in the coherent superposition $\psi = \varphi + \varphi_0$, that cannot be distinguished by the measurement of $|\varphi_0(x)|^2 = |\psi(x)|^2$ ($c_+ = 0$ since $c(x) = 0$ for all $x \in \Omega$). At the same time, such wave packets are easily distinguished in the momentum representation:

$$\langle\tilde{C}\rangle_+ = |\tilde{v}|^{-1} \int (e^{-\pi|(x-\eta)v^{-1}|^2} - e^{-\pi|x\tilde{v}^{-1}|^2})\, dx \simeq 1$$

if $|\eta v^{-1}| \gg 1$ since in this case $\tilde{c}_+(x) \simeq |\tilde{\psi}_{0\eta}(x)|^2$.

In general, for every quasiselective measurement of intensity on a Borel space X with a positive operator measure $M(\Delta) \leqslant I$, $\Delta \subseteq X$, optimal detection is determined by the solution to the problem

$$\varkappa_M^0(C) = \sup_{\Delta \subseteq X} \langle C, M(\Delta)\rangle = \varkappa(\Delta^0) \qquad (2.2.1.14)$$

of finding the upper bound of the degree-of-contrast measure $\varkappa(\Delta) = \langle C, M(\Delta)\rangle$. The supremum (2.2.1.14) is attained on the $|\varkappa|$-measurable set Δ^0, the support of the positive part $\varkappa_+ = 0 \vee \varkappa = (\varkappa + |\varkappa|)/2$ of measure $\varkappa$:

$$\Delta^0 = \bigcap \{\bar{\Delta}: \varkappa_+(\Delta) = 0\} \equiv \Delta_+, \qquad (2.2.1.15)$$

which realizes the lower bound in the positive measures $\lambda \geqslant \varkappa$ of finite variation:

$$\langle\varkappa\rangle_+ = \inf_{\lambda \geqslant 0} \{\lambda(X) \mid \lambda \geqslant \varkappa\} = \varkappa_+(X), \qquad (2.2.1.16)$$

which determines the gauge $\langle\varkappa\rangle_+ = 0 \Leftrightarrow \varkappa \leqslant 0$ of measure $\varkappa$.

2.2.1.3 Optimal Detection

As the example discussed in Section 2.2.1.2 shows, the quality of detection, which is determined for a given intensity distribution on X by the degree-of-contrast measure $\mu(C, \Delta) = \langle C, M(\Delta)\rangle$, must be optimized not only with respect to measurement regions $\Delta \subseteq X$ but also with respect to the methods of measurement of this quantity. These methods are determined by the ways in which the positive operator-valued measure $M(\Delta) \leqslant E$ is specified, where E is any orthoprojector in $\mathscr{H}$ satisfying the condition $CE = C$. Here it is sufficient to find at least one resolving operator $D = M(\Delta)$ that realizes the upper bound of the maximal degree of contrast (2.2.1.14):

$$\varkappa^0(C) = \sup_{D \geqslant 0} \{\langle C, D\rangle \mid D \leqslant E\}. \qquad (2.2.1.17)$$

Employing the methods of linear programming in partially ordered Banach spaces [2.48], we can formulate the necessary and sufficient conditions for the optimality of the detection operator D employing

criterion (2.2.1.17), which is determined by the trace class degree-of-contrast operator $C = R - P_0$.

Theorem 2.1.1 *The upper bound* (2.2.1.17) *is attained on operator* $0 \leqslant D^0 \leqslant E$ *if and only if*

$$B^0 (E - D^0) = 0, \quad (B^0 - C) D^0 = 0, \qquad (2.2.1.18)$$

where $B^0 \geqslant 0, C$. *The operator* B^0 *here is the solution to the duality problem*

$$\langle C \rangle_+ = \inf_{B \geqslant 0} \{\langle B, E\rangle \mid B \geqslant C\} \qquad (2.2.1.19)$$

for which the conditions (2.2.1.18) *for admissible* D^0 *are also necessary and sufficient, with* $\varkappa^0 (C) = \langle C \rangle_+$.

Proof. The sufficiency of the optimality conditions (2.2.1.18) for solving problems (2.2.1.17) and (2.2.1.19) can be verified directly by employing the property of the monotonicity of the trace, $B \geqslant C \Rightarrow$ $\mathrm{Tr}\,(BD) \geqslant \mathrm{Tr}\,(CD)$, for every positive operator D. Allowing for the fact that $B^0E = B^0D^0 = CD^0$ for every $0 \leqslant D \leqslant E$, we obtain

$$\langle C, D \rangle = \mathrm{Tr}\,(CD) \leqslant \mathrm{Tr}\,(B^0D) \leqslant \mathrm{Tr}\,(B^0E) = \langle C, D^0 \rangle.$$

Similarly, for every $B \geqslant 0$ and every C we obtain

$$\langle B, E \rangle = \mathrm{Tr}\,(BE) \geqslant \mathrm{Tr}\,(BD^0) \geqslant \mathrm{Tr}\,(CD)^0 = \langle B^0, E \rangle.$$

The necessity of the optimality conditions (2.2.1.18) follows from the fact that the inequality

$$\langle C, D \rangle = \mathrm{Tr}\,(CD) \leqslant \mathrm{Tr}\,(BD) \leqslant \mathrm{Tr}\,(BE) = \langle B, E \rangle, \qquad (2.2.1.20)$$

which is valid for all operators D and B admissible in problems (2.2.1.17) and (2.2.1.19), must transform into the equality $\langle C, D^0 \rangle = \langle B^0, E \rangle$ on the extremal operators D^0 and B^0, in accordance with Lagrange's principle of duality:

$$\begin{aligned} \sup_{D \geqslant 0} \{\langle C, D\rangle \mid D \leqslant E\} &= \sup_{D \geqslant 0} \inf_{B \geqslant 0} \{\langle C, D\rangle + \langle B, E - D\rangle\} \\ &= \inf_{B \geqslant 0} \sup_{D \geqslant 0} \{\langle C - B, D\rangle + \langle B, E\rangle\} \\ &= \inf_{B \geqslant 0} \{\langle B, E\rangle \mid B \geqslant C\}. \end{aligned}$$

Whereby, allowing for the fact that the trace of the product of positive operators is zero if and only if the product itself is zero, we arrive at conditions (2.2.1.18) via the following relation:

$$\mathrm{Tr}\,[B^0 (E - D^0)] + \mathrm{Tr}\,[(B^0 - C) D^0] = \mathrm{Tr}\,(B^0E) - \mathrm{Tr}\,(CD^0) = 0.$$

The proof of the theorem is complete.

Note that the solutions to problems (2.2.1.17) and (2.2.1.19) exist for every Hermitian trace class operator C and every bounded

positive orthoprojector E; the solution to problem (2.2.1.17) is unique only if E is the minimal of the orthoprojectors for which $CE = C$, while the solution to problem (2.2.1.19) is unique only if E is the maximal $E = I$ of the orthoprojector E. Indeed, employing the spectral representation of operator C, we write this operator in the form of the orthogonal sum

$$C = \sum \varkappa_n \mid \chi_n)(\chi_n \mid = C_+ + C_- \tag{2.2.1.21}$$

of the positive and negative operators

$$C_+ = \sum_{\varkappa_n > 0} \varkappa_n \mid \chi_n)(\chi_n \mid, \quad C_- = \sum_{\varkappa_n < 0} \varkappa_n \mid \chi_n)(\chi_n \mid, \tag{2.2.1.22}$$

where we have allowed for the fact that a Hermitian trace class operator has a discrete spectrum of finite multiplicity, $\varkappa_n \in \mathfrak{R}$, which can be found by solving the eigenvalue problem $C\chi = \varkappa\chi$. The orthoprojector E satisfying condition $CE = C$ can be written in the form of the orthogonal sum

$$E = E_+ + E_0 + E_-, \tag{2.2.1.23}$$

where $E_+ = \sum_{\varkappa_n > 0} \mid \chi_n)(\chi_n \mid$, $E_- = \sum_{\varkappa_n > 0} \mid \chi_n)(\chi_n \mid$, and $E_0 = E - E_+ - E_-$. The operators $D^0 = E_+$, $B^0 = C_+$ are, obviously, admissible: $0 \leqslant E_+ \leqslant E$, $C_+ \geqslant 0$, C and optimal:

$$\begin{gathered} C_+ (E - E_+) = C_+ (E_0 + E_-) = 0, \\ (C_+ - C) E_+ = -C_- E_+ = 0. \end{gathered} \tag{2.2.1.24}$$

Every other solution D^0 to problem (2.2.1.17) satisfies conditions (2.2.1.18) for $B^0 = C_+$:

$$\begin{aligned} C_+ (E - D^0) &= C_+ - C_+ D^0 = 0, \\ (C_+ - C) D^0 &= -C_- D^0 = 0, \end{aligned}$$

in view of which $E_+ = E_+ D^0$ and $E_- D^0 = 0$, that is,

$$E_+ \leqslant D^0 \leqslant E - E_- = E_+ + E_0. \tag{2.2.1.25}$$

Similarly, every solution B^0 to problem (2.2.1.19) satisfies conditions (2.2.1.18) for $D^0 = E_+$:

$$B^0 (E - E_+) = 0, \quad (B^0 - C) E_+ = B^0 E_+ - C_+ = 0,$$

which imply that B is commutative with E_+ and, hence can be represented in the form of the orthogonal sum $B^0 = B_+ + B_0$, with

$$B_+ = B^0 E_+ = C_+, \quad B_0 (E - E_+) = 0,$$

that is,

$$B^0 = C_+ + B_0, \quad B_0 \geqslant 0, \quad B_0 E = 0. \tag{2.2.1.26}$$

Thus, the general solution to the problem of optimal detection is determined by the quasifilter (2.2.1.25) of the form $D^0 = E_+ + D_0$, where D_0 is an arbitrary operator, $0 \leqslant D_0 \leqslant E_0$, and an ideal filter $D^0 = E_+$ if $E = E_+ + E_-$. The general solution to the duality problem (2.2.1.19) is determined by the operator of the form (2.2.1.26), with $B_0 = 0$ at $E = I$. The maximal possible degree of contrast realized by the optimal detector D^0 is given by the expression

$$\varkappa_+ (R - P_0) = \mathrm{Tr}\ (R - P_0)_+ = \sum_{\varkappa_n > 0} \varkappa_n. \tag{2.2.1.27}$$

2.2.1.4 Coherent and Quasioptimal Detection

Let us consider the particular problem of optimal detection of a wave pattern described by a common amplitude $\varphi \in \mathscr{H}$ in a partially coherent mixture with a generalized random amplitude H_0: $\mathscr{H} \to \mathscr{K}$. The resulting amplitude $G = |\,\xi)\,(\varphi\,| + H_0$, with $\xi \in \mathscr{K}$ a normalized vector $\|\,\xi\,\| = 1$, defines a density operator $R = G^+G$ of the form

$$R = |\,\varphi)\,(\varphi\,| + |\,\varphi)\,(\varphi_0\,| + |\,\varphi_0)\,(\varphi\,| + P_0, \tag{2.2.1.28}$$

with $\varphi_0 = F_0^+\xi \in \mathscr{H}$ and $P_0 = F_0^+F_0$ the background-density operator. Thus, we are required to solve the extremal problem (2.2.1.17) for the two-dimensional degree-of-contrast operator $C = R - P_0$ of the form

$$\begin{aligned} C &= |\,\varphi)\,(\varphi\,| + |\,\varphi)\,(\varphi_0\,| + |\,\varphi_0)\,(\varphi\,| \\ &= |\,\psi)\,(\psi\,| - |\,\varphi_0)\,(\varphi_0\,|, \end{aligned}$$

which corresponds to the coherent superposition $\psi = \varphi + \varphi_0$ of the common amplitudes φ and φ_0. We will consider this problem in the minimal subspace $\mathscr{E} \subset \mathscr{H}$ generated by the amplitudes $\psi_0 = \varphi_0$ and $\psi_1 = \varphi_0 + \varphi$. For its solution we find the eigenvectors and eigenvalues of operator C by constructing the secular equation $C\chi = \varkappa\chi$ for the coefficients of the expansion $\chi = \alpha_0\psi_0 + \alpha_1\psi_1$ in the base $\{\psi_0, \psi_1\}$ of space $\mathscr{E}$:

$$\psi_1\,(\psi_1\,|\,\alpha_0\psi_0 + \alpha_1\psi_1) - \psi_0\,(\psi_0\,|\,\alpha_0\psi_0 + \alpha_1\psi_1) = \varkappa\,(\alpha_0\psi_0 + \alpha_1\psi_1). \tag{2.2.1.29}$$

Introducing the notation $\nu_i = \|\,\psi_i\,\|^2$, $i = 0, 1$, $\beta = (\psi_0\,|\,\psi_1)$, and equating the coefficients of ψ_i, $i = 0, 1$, in Eq. (2.2.1.29), we arrive at a system of two homogeneous equations,

$$(\nu_0 + \varkappa)\,\alpha_0 + \beta\alpha_1 = 0, \qquad \beta\alpha_0 + (\nu_1 - \varkappa)\,\alpha_1 = 0. \tag{2.2.1.30}$$

This system has nonzero solutions only if the system determinant is zero, or

$$(\nu_0 + \varkappa)(\nu_1 - \varkappa) - |\beta|^2 = 0. \tag{2.2.1.31}$$

Solving this quadratic equation for $\varkappa$, we obtain the eigenvalues:

$$\varkappa_{\pm} = \frac{\nu_1 - \nu_2}{2} \pm \sqrt{\left(\frac{\nu_1 + \nu_0}{2}\right)^2 - |\beta|^2}, \tag{2.2.1.32}$$

which are real in view of the Schwarz inequality

$$|\beta|^2 = |(\psi_0 | \psi_1)|^2 \leqslant |\psi_0|^2 \, |\psi_1|^2 = \nu_0 \nu_1,$$

and, obviously, have opposite signs: $\pm\varkappa_{\pm} \geqslant 0$. At $\beta = 0$ the amplitudes ψ_0 and ψ_1 are orthogonal, $\varkappa_+ = \nu_1$, $\varkappa_- = \nu_0$, and the eigenvectors φ_+ and φ_- coincide with the normalized amplitudes $\psi_1/\sqrt{\nu_1}$ and $\psi_0/\sqrt{\nu_0}$, respectively. Optimal detection in this case is reduced to the discrimination of amplitudes ψ_1 and ψ_0 by measuring the degree of contrast of oscillations in the resulting mode $\chi_+ = \psi_1/\sqrt{\nu_1}$, which is equal to the intensity of oscillations at $\varkappa = \nu_1$ in this mode if the received signal is ψ_1 and to zero if the received signal is ψ_0. In the opposite case $|\beta|^2 = \nu_0\nu_1$ of the colinearity of ψ_1 and ψ_0, the values $\varkappa_{\pm}$ are equal, respectively, to the positive and negative parts of the difference $\nu_1 - \nu_0$:

$$\varkappa_{\pm} = \frac{1}{2}(\nu_1 - \nu_0 \pm |\nu_1 - \nu_0|) = (\nu_1 - \nu_0)_{\pm}.$$

The corresponding optimal detection is reduced to the measurement of the positive degree of contrast $\varkappa_+ = \nu_1 - \nu_0$ in the mode $\chi = \psi_1/\sqrt{\nu_1} = \psi_0/\sqrt{\nu_0}$ if $\nu_1 > \nu_0$, in the opposite case, $\nu_0 \geqslant \nu_1$, the degree of contrast $\varkappa_+$ is zero and no measurement is carried out, or $\chi_+ = 0$. The optimal detection of a wave pattern φ of intensity $\mu = \|\varphi\|^2 \neq 0$ in the coherent superposition $\psi = \varphi + \varphi_0$ is therefore reduced to the measurement of the maximal degree of contrast

$$\varkappa_+ = \sqrt{\mu\nu_0}\left(\operatorname{Re}\gamma + \sqrt{\lambda} + \sqrt{(\operatorname{Re}\gamma + \sqrt{\lambda})^2 + 1 - |\gamma|^2}\right), \tag{2.2.1.33}$$

where $\gamma = (\varphi_0 | \varphi)\sqrt{\mu\nu_0}$ is the coefficient of colinearity of amplitudes φ and φ_0, and λ is the signal-to-noise ratio. The corresponding ideal filter $E_+ = |\chi_+)(\chi_+|$ is defined at $\varkappa_+ \neq 0$ by the eigenvector $\chi_+ = \alpha_+\varphi + \alpha_0\varphi_0$ with coefficient

$$\alpha_+ = \sqrt{\nu_0\mu}\,\alpha_0\left(j\operatorname{Im}\gamma + \sqrt{\lambda} + \sqrt{(\operatorname{Re}\gamma + \sqrt{\lambda})^2 + 1 - |\gamma|^2}\right), \tag{2.2.1.34}$$

$\alpha_0 > 0$, found from the normalization condition $\| \chi_+ \| = 1$. The case where $\varkappa_- = 0$, and therefore $\chi_+ = 0$, is possible in the minimal subspace $\mathscr{E}$ only if φ and φ_0 are colinear, when $|\gamma| = 1$, and

$$\varkappa_+ = \sqrt{\mu\nu_0}(\cos\theta + \sqrt{\lambda})_+, \tag{2.2.1.35}$$

where $\cos\theta = \operatorname{Re}\gamma$. The optimal filter in this case is matched with the signal mode $\chi_+ = \varphi/\sqrt{\mu}$ if $\cos\theta > -\sqrt{\lambda}$, and $\chi = 0$ in the opposite case if $\cos\theta \leqslant -\sqrt{\lambda}$ which is possible only if $\lambda \leqslant 1$.

The same filter $\chi_0 = \varphi/\sqrt{\mu}$ matched with ψ is used to describe the asymptotically optimal detection at large signal-to-noise ratios $\lambda = 1/\varepsilon \gg (\operatorname{Re}\gamma)^2$. The degree of contrast is then

$$\varkappa_0 = \mu(1 + \sqrt{\varepsilon}\operatorname{Re}\gamma), \tag{2.2.1.36}$$

which coincides with (2.2.1.33) to within ε. In the next order we obtain a filter matched with the resulting mode $\chi_1 = \psi/\sqrt{\nu_1}$ and realizing the degree of contrast

$$\varkappa_1 = \mu\left(1 + \frac{\sqrt{\varepsilon}}{2}\operatorname{Re}\gamma + \frac{\varepsilon}{4} - \frac{\varepsilon\,|\sqrt{\varepsilon}/2 + \gamma|^2}{4[1 + (\sqrt{\varepsilon}/2)\operatorname{Re}\gamma + \varepsilon/4]}\right). \tag{2.2.1.37}$$

For an orthogonal background, $\varphi \perp \varphi_0$, we have $\gamma = 0$, and the normalized eigenvector χ_+ corresponding to the eigenvalue

$$\varkappa_+ = \sqrt{\mu\nu_0}(\sqrt{1+\lambda} + \sqrt{\lambda}) \tag{2.2.1.38}$$

can be written in the form

$$\chi_+ = (\varphi_0/\sqrt{\nu_0} + (\sqrt{1+\lambda} + \sqrt{\lambda})\,\varphi/\sqrt{\mu})/(2(1+\lambda) + \sqrt{\lambda(1+\lambda)}).$$

For $\operatorname{Im}\gamma \neq 1$ and a low signal-to-noise ratio $\lambda \ll 1 - (\operatorname{Im}\gamma)^2$, the maximal degree of contrast is realized at

$$\chi_+ = (1/\sqrt{2})(e^{j\theta}\varphi/\sqrt{\mu} + \varphi_0/\sqrt{\nu_0}), \tag{2.2.1.39}$$

with $\sin\theta = \operatorname{Im}\gamma$, and is determined asymptotically by the expression

$$\varkappa_+ \approx \sqrt{\mu\nu_0}(\operatorname{Re}\gamma + \cos\theta + \sqrt{\lambda}(\operatorname{Re}\gamma + 1)), \tag{2.2.1.40}$$

with $\cos\theta = \sqrt{1-(\operatorname{Im}\gamma)^2}$. For one, at $\gamma = 0$ we get $\varkappa_+ \approx \sqrt{\mu\nu_0} \times (1 + \sqrt{\lambda})$.

In the general case of the partially coherent superposition $G = F + F_0$, the solution of the eigenvalue problem for the degree-of-contrast operator $C = G^+G - F_0^+F_0$ constitutes a complicated mathematical problem. If we isolate the coherent component from the

generalized amplitude F_0, we can represent the latter in the form

$$F_0 = \frac{1}{2}\sqrt{\varepsilon}FA^+ + W,$$

with W the incoherent component, $F^+W = 0$, and A is an operator in $\mathcal{H}$, which we assume to be bounded: $\|A\| \leqslant 1$. Next we select the positive constant ε in an appropriate manner. Operator C then assumes the form

$$C = P + \frac{1}{2}\sqrt{\varepsilon}(PA^+ + AP) \tag{2.2.1.41}$$

and is a trace class operator if $P = F^+F$ is an operator with a finite trace.

For high signal-to-noise ratios $\lambda = 1/\varepsilon \gg 1$, the eigenvectors χ_n and the corresponding eigenvalues $\varkappa_n$ of operator C can be found via perturbation theory methods. In the first order in $\sqrt{\varepsilon}$ the eigenvectors coincide with the eigenvectors φ_n of the signal density operator P, that is, $P\varphi_{0n} = \mu_n\varphi_{0n}$, and realize the following degrees of contrast:

$$\varkappa_{0n} = (\varphi_{0n} \mid C\varphi_{0n}) = \mu_n(1 + \sqrt{\varepsilon}\operatorname{Re}\gamma_n), \tag{2.2.1.42}$$

with $\gamma_n = (A\varphi_n \mid \varphi_n)$. The corresponding quasioptimal detection is reduced, therefore, to measuring the total degree of contrast

$$\varkappa_0 = \operatorname{Tr}(CE_0) = \sum_{n\in N_+} \mu_n(1 + \sqrt{\varepsilon}\operatorname{Re}\gamma_n) \tag{2.2.1.43}$$

matched with the signal orthogonal modes φ_n of the ideal filter

$$E_0 = \sum_{n\in N_+} |\varphi_n)(\varphi_n|, \quad N_+ = \{n\colon \operatorname{Re}\gamma_n > -1/\sqrt{\varepsilon}\}. \tag{2.2.1.44}$$

When the intensities of the signal and the noise are comparable, $\varepsilon \approx 1$, the quality of such detection may be considerably lower than that of optimal detection. In particular, for the above example of orthogonal φ and φ_0 we have $\gamma = 0$ and $\varkappa_0 = \mu$, while the quality of optimal detection (2.2.1.38) equal to $\varkappa_+ = \mu(1 + \sqrt{1+\varepsilon})/2 > \mu$ is more than twice as great as $\varkappa_0$ if the signal intensity is less than half of the intensity of the noise, and we have

$$\varkappa_+/\varkappa_0 = (1 + \sqrt{1+\varepsilon})/2 \to \infty \text{ as } \varepsilon = \nu_0/(4\mu) \to \infty. \tag{2.2.1.45}$$

2.2.2 Multialternative Detection and Identification of Wave Patterns

In this section we will consider the problem of detecting one of several simple or mixed wave patterns that is in a partially coherent superposition with the background. We will introduce the necessary and sufficient conditions for the optimality of such detection, using the criterion of the maximum

degree of contrast, and give these conditions a concrete meaning for the problem of separating such patterns from an incoherent background. We will also give the complete solution to the problem of identifying nonorthogonal waves of the same intensity. This solution coincides with that of the problem of optimal discrimination between pure quantum states obtained in [2.4, 2.5]. Finally, we will discuss the quasioptimal method of multialternative detection based on the perturbation theory for degree-of-contrast operators.

2.2.2.1 Statement of the Problem

The problem of m-alternative detection of sound and visual patterns described in a given spatial-frequency region Ω by the wave amplitudes $\varphi_i\ (q)$, $i = 1, \ldots, m$ belonging to the Hilbert space $\mathscr{H} = \mathscr{L}^2\ (\Omega)$ can be solved in a trivial manner in the absence of an audio or optical background, $\varphi_0 = 0$, only on the assumption that these amplitudes are pairwise orthogonal, $(\varphi_i \mid \varphi_k) = 0$ for $i \neq k$. It is sufficient to measure the intensity distribution $\varepsilon_i = \mid (\psi \mid \chi) \mid^2$ in the received signal $\psi \in \{\varphi_i\}_{i=1}^m$ over the orthogonal modes $\chi_k = \varphi_k/\| \varphi_k \|$, $i = 1, \ldots, m$, to determine correctly the pattern $\varphi = \varphi_i$ with a nonzero intensity $\mu_i = \| \varphi_i \|^2$ by specifying the number of the excited mode i: $\varepsilon_i = \mu_i \neq 0$. The other modes χ_k, $k \neq i$, remain unexcited in the process, and the case $\varepsilon_i = 0$ for all $i = 1, \ldots, m$ means that these patterns are absent from the measurement region Ω.

The simplest problem of multialternative detection in noise, the problem of isolating one of a set of orthogonal amplitudes $\{\varphi_i\}$ from an incoherent mixture $R_i = \mid \varphi_i)\ (\varphi_i \mid + N$ with an optical or acoustic background not necessarily described by a trace class density operator N, has the same solution if we compare the intensities $\varepsilon_i = (\chi_i \mid R\chi_i) = \mu_i + \nu_i$ of the received signal $R \in \{R_i\}_{i=0}^m$ not with zero but with the background level $\nu_i = (\chi_i \mid N\chi_i)$ in the orthogonal modes $\chi_i = \varphi_i/\| \varphi_i \|$, $i = 1, \ldots, m$.

In the case of nonorthogonal amplitudes $\{\varphi_i\}_{i=1}^m$ there is no way of measuring the intensity distribution in the received signal directly over the modes $\varphi_i/\| \varphi_i \|$. We must therefore find a set $\{\chi_i\}_{i=1}^m \subset \mathscr{H}$ that satisfies the condition $\sum\limits_{i=1}^{m} \mid \chi_i)\ (\chi_i \mid \leqslant I$ and for which a pattern φ_i can be confidently reconstructed from the distribution $\varkappa_i = \mid (\psi \mid \chi_i) \mid^2$ corresponding to the received signal $\psi \in \{\varphi_i\}_{i=1}^m$.

The same problem emerges in the multialternative detection of patterns $\{\varphi_i\}_{i=1}^m$ in the coherent superposition $\psi_i = \varphi_i + \varphi_0$ with a nonzero background amplitude φ_0 even when all the amplitudes $\{\varphi_i\}_{i=1}^m$ are mutually orthogonal and orthogonal to φ_0. Although in the latter case we can still measure the intensities in the orthogonal signal modes $\chi_i = \varphi_i/\| \varphi_i \|$ and this yields a total degree of contrast $\varkappa_0 = \sum\limits_{i=1}^{m} \mu_i$, we can achieve a higher quality of detection

if we minimize the expression

$$\varkappa = \sum_{i=1}^{m} (\chi_i | C_i \chi_i) = \sum_{i=1}^{m} |(\chi_i | \psi_i)|^2 - |(\chi_i | \varphi_0)|^2, \tag{2.2.2.1}$$

where

$$\begin{aligned} C_i &= |\varphi_i)(\varphi_i| + |\varphi_0)(\varphi_i| + |\varphi_i)(\varphi_0| \\ &= |\psi_i)(\psi_i| - |\varphi_0)(\varphi_0|, \end{aligned} \tag{2.2.2.2}$$

as we did in the case with $m = 1$ in the example at the end of Section 2.2.1.

Note that in the coordinate representation the wave patterns $\{\varphi_i\}_{i=1}^m$ may be indistinguishable even if they are orthogonal, as is the case, say, for harmonic amplitudes $\varphi_i(f) = \exp\{2\pi jfi/\Phi\}$ which are orthogonal in the frequency interval $[0, \Phi]$ and have the same homogeneous distributions, $|\varphi_i(f)|^2 = 1$. The maximal quality of m-alternative detection achievable through measurements of the coordinate distribution of the degree of contrast

$$c_i(x) = |\varphi_i(x)|^2 + 2\mathrm{Re}\, \varphi_i^*(x)\varphi_0(x) = |\psi_i(x)|^2 - |\varphi_0(x)|^2 \tag{2.2.2.3}$$

is determined by the solution to the extremal problem

$$\varkappa_I^0(C) = \sup_{\sum_{i=1}^m \Delta_i \subseteq \Omega} \sum_{i=1}^m \langle C_i, I(\Delta_i)\rangle = \sum_{i=1}^m \int_{\Delta_i^0} c_i(x)\, dx, \tag{2.2.2.4}$$

where the supremum is taken over the measurable nonintersecting subsets $\Delta_i \subset \Omega$ of a coordinate region Ω that can be bound by the union of the supports of the integrable functions $c_i(x)$, $i = 1, \ldots, m$. This limit is attained, obviously, on the partitions $\Omega_+ = \sum_{i=1}^m \Delta_i^0$ of the measurable set Ω_+ in every point of which at least one of the functions $c_i(x)$ is positive and coincides on Δ_c^0 with the upper envelope $c_\vee(x) = \max_{i=1,\ldots,m} c_i(x)$. Thus, the total degree of contrast (2.2.2.4) coincides with the integral over Ω of the positive part $c_+(x) = \max(0, c_\vee(x))$ of $c_\vee$, which determines the solution of the duality problem

$$\langle \mathbf{c}\rangle_+ = \inf_{b \geqslant 0}\left\{\int_\Omega b(x)\,dx \,|\, b \geqslant c_i,\ i = 1, \ldots, m\right\} = \int_\Omega c_+(x)\,dx. \tag{2.2.2.5}$$

The lower bound (2.2.2.5) over all positive integrable functions $b(x) \geqslant 0$, which almost everywhere majorize every function $c_i(x)$, defines the positive gauge of the vector function $\mathbf{c} = \{c_i\}_{i=1}^m$, $\langle \mathbf{c}\rangle_+ =$

$0 \Leftrightarrow c_i(x) \leqslant 0$. The best m-alternative detection in the coordinate region Ω is reduced, therefore, to a search among the Δ_i for the regions Δ_i^0 on which the measured degree of contrast $\mathbf{c}(x)$ is positive and reaches $\mathbf{c}_+(x)$; in the opposite case of $\mathbf{c}(x) < \mathbf{c}_+(x)$ the pattern may not be detected for all $x \in \Omega_+$ and it can be assumed to be undetected if $\mathbf{c}(x) \leqslant 0$ for all $x \in \Omega$.

2.2.2.2 The Optimality Conditions

Let us consider the problem of maximizing the quality of m-alternative detection of mixed patterns F_i: $\mathscr{H} \to \mathscr{K}$ with trace class density operators $P_i = F_i^\dagger F_i$, $i = 1, \ldots, m$, in a partially coherent superposition $G_i = F_i + F_0$ for which the mutual densities $F_i^\dagger F_0 = \sqrt{\varepsilon} P_i A_i^\dagger/2$ are determined by the contraction operators A_i: $\mathscr{H} \to \mathscr{H}$ ($\varepsilon > 0$ is a parameter). The respective extremal problem is formulated for trace class operators of the degree of contrast,

$$C_i = R_i - P_0 = P_i + \frac{1}{2}\sqrt{\varepsilon}\,(P_i A_i^\dagger + A_i P_i), \quad i = 1, \ldots, m, \tag{2.2.2.6}$$

$R_i = G_i^\dagger G_i$, in the class of quasiselective measurements described by any resolving operators $\{D_i\}_{i=1}^m$:

$$\varkappa^0(\mathbf{C}) = \sup_{D_i \geqslant 0} \left\{ \sum_{i=1}^m \langle C_i, D_i \rangle \,\middle|\, \sum_{i=1}^m D_i \leqslant E \right\}, \tag{2.2.2.7}$$

where E is an operator for which $C_i E = C_i$ for all $i = 1, \ldots, m$.

Theorem 2.2.2.1 *The upper bound* (2.2.2.7) *is attained on the admissible operators D_i, $i = 1, \ldots, m$, if and only if there is a trace class operator $B^0 \geqslant 0$, C_i, $i = 1, \ldots, m$, such that*

$$B^0(E - D^0) = 0, \; (B^0 - C_i) D_i^0 = 0, \; i = 1, \ldots, m, \tag{2.2.2.8}$$

with $D^0 = \sum_{i=1}^m D_i^0$. The operator B^0 then is the solution to the duality problem

$$\langle \mathbf{C} \rangle_+ = \inf_{B \geqslant 0} \{ \langle B, E \rangle \,|\, B \geqslant C_i, \; i = 1, \ldots, m \}, \tag{2.2.2.9}$$

for which conditions (2.2.2.8) *are also necessary and sufficient when $D_i^0 \geqslant 0$, $\sum_{i=1}^m D_i^0 \leqslant E$, with $\varkappa^0(\mathbf{C}) = \langle \mathbf{C} \rangle_+$.*

Proof. The proof, which is similar to the proof of a particular case of this theorem, Theorem 2.2.1.1, will be found as a corollary of a more general theorem, Theorem 2.2.3.1.

It can also be easily proved that the solution to problem (2.2.2.7) exists for all trace class operators C_i and determines on subspace

$\mathscr{E} = E\mathscr{H}$ for which $C_i E = C_i$ for all $i = 1, \ldots, m$, a unique solution $B^0 = B^0 E$ to problem (2.2.2.9).

Indeed, the lower bound (2.2.2.9) determines for the vector operators $\mathbf{C} = \{C_i\}_{i=1}^m$ the gauge $\langle \mathbf{C} \rangle_+$, a positive homogeneous sublinear functional on the space of families $\{C_i\}_{i=1}^m$ of the trace class operators C_i, $i = 1, \ldots, m$, that possesses the property $\langle \mathbf{C} \rangle_+ = 0 \Leftrightarrow C_i \leqslant 0$ for all i's. Bearing in mind that every linear functional $\mathbf{C} \mapsto \langle \mathbf{C}, \mathbf{D} \rangle$ satisfying the condition $\langle \mathbf{C}, \mathbf{D} \rangle \leqslant \langle \mathbf{C} \rangle_+$ is positive and has the form $\langle \mathbf{C}, \mathbf{D} \rangle = \sum_{i=1}^m \mathrm{Tr}\,(C_i, D_i)$, where $\sum_{i=1}^m D_i = E$, we find that the set that is conjugate to $\{\mathbf{C} \mid \langle \mathbf{C} \rangle_+ \leqslant 1\}$ consists of the resolving families $\mathbf{D} = \{D_i\}_{i=1}^m$ of bound operators that are admissible in problem (2.2.2.7). The existence of a solution to problem (2.2.2.7) follows, therefore, from the Hahn-Banach theorem, according to which for every vector $\mathbf{C}^0$ of a calibrated space there exists a supporting functional $\mathbf{D}^0$ defined by the conditions $\langle \mathbf{C}, \mathbf{D}^0 \rangle \leqslant \langle \mathbf{C} \rangle_+$ and $\langle \mathbf{C}^0, \mathbf{D}^0 \rangle = \langle \mathbf{C}^0 \rangle_+$. For every solution $\mathbf{D}^0$ to problem (2.2.2.7) the solution of the conjugate problem (2.2.2.9) on the subspace $\mathscr{E} = E\mathscr{H}$ is determined uniquely by the formula

$$B^0 E = B^0 D^0 = \sum_{i=1}^m C_i D_i^0 \tag{2.2.2.10}$$

which is obtained by adding (2.2.2.8) over $i = 1, \ldots, m$. Note that the above proof of the existence of a solution to problem (2.2.2.7) and of the uniqueness of the solution to problem (2.2.2.9) remains valid for the case of an infinite number of patterns $m = \infty$ if we require that $\langle C_i, E \rangle = \mathrm{Tr}\, C_i \to 0$ as $i \to \infty$.

Conditions (2.2.2.8) can easily be met for $m > 1$ by analogy with the case of $m = 1$ only for commutative C_i, when these operators have a joint spectral representation

$$C_i = \sum_{i=1}^m \varkappa_{in} |\chi_n)(\chi_n|, \quad (\chi_n|\chi_m) = \delta_{nm}. \tag{2.2.2.11}$$

The orthoprojector E can be resolved into an orthogonal sum $E = E_0 + \sum_{i=1}^m E_i$, where

$$E_i = \sum_{n \in \mathbb{N}_i} |\chi_n)(\chi_n|, \quad \mathbb{N}_i \subseteq \{n \in \mathbb{N} \mid \varkappa_{in} \geqslant 0, \ \varkappa_{kn}, \ k \neq i\}$$

(points n at whict $\varkappa_{in} = \max_{j=1,\ldots,m} \varkappa_{jn} = \varkappa_{kn}$ refer to any one of the nonintersecting sets $\mathbb{N}_i$, $\mathbb{N}_h$). The operators

$$D_i^0 = E_i, \quad B^0 = \sum_{i=1}^m \sum_{n \in \mathbb{N}_i} \varkappa_{in} |\chi_n)(\chi_n| = \sum_{i=1}^m C_i E_i \tag{2.2.2.12}$$

are, therefore, admissible and optimal:

$$B^0 (E - D^0) = B^0 E_0 = 0, \quad (B^0 - C_i) E_i = C_i E_i - C_i E_i = 0.$$

Thus, optimal m-alternative detection in the commutative case is reduced to measuring the discrete distribution of the degree of contrast $\varkappa$ in the proper representation of operators C_i. The total maximal degree of contrast in this case is determined from the formula

$$\varkappa^0 (\mathbf{C}) = \sum_{i=1}^{m} \sum_{n \in \mathbb{N}_i} \varkappa_{in} = \sum_{n \in \mathbb{N}} \max_i \{\varkappa_{in} \vee 0\}. \tag{2.2.2.13}$$

2.2.2.3 Optimal Identification

Let us consider the important case of positive operators $C_i = H_i^+ H_i = S_i$ which occur, say, in the case of an incoherent superposition of wave patterns F_i with a background F_0, when the degrees of contrast (2.2.2.6) are the density operators $P_i = F_i^+ F_i$. The corresponding extremal problem (2.2.2.7) of pattern recognition, which is known as the optimal identification problem, is not trivial for noncommutative S_i, $i = 1, \ldots, m$, for $m > 1$ even if these patterns are pure, that is, are described by nonorthogonal amplitudes ψ_i, $i = 1, \ldots, m$. For the case of $m = 2$, however, the optimal identification problem can easily be solved by reducing it to the problem of optimal detection with one degree-of-contrast operator $C = S_1 - S_2$. Indeed, allowing for the fact that the admissible operators $B = L$ in the duality problem (2.2.2.9) are determined by the conditions $L \geqslant B > 0$, $i = 1, 2$, we can proceed from (2.2.2.9) to (2.2.1.19) by carrying out the substitution inf $\langle L, E \rangle = \langle S_2, E \rangle +$ inf $\langle B, E \rangle$ where $B = L - S_2$ is the admissible operator of problem (2.2.1.19): $B \geqslant \{S_1 - S_2, S_2 - S_2\} = \{C, 0\}$. Thus, the solution D^0 to problem (2.2.1.17) makes it possible to represent the solution to problem (2.2.2.7) in the form

$$\varkappa^0 (\mathbf{S}) = \langle C, D^0 \rangle + \langle S_2, E \rangle = \langle S_1, D^0 \rangle + \langle S_2, E - D^0 \rangle,$$

which yields the optimal decision operators $D_1^0 = D^0$ and $D_2^0 = E - D^0$.

To investigate the problem of identifying wave patterns in the multialternative case with $m > 2$, we restrict the space $\mathscr{H}$ by the minimal space $\mathscr{E}^0 \subseteq \mathscr{H}$ containing all the ranges of values $\mathscr{H}_i = S_i \mathscr{H}$. Since the $S_i \geqslant 0$, every operator B admissible to problem (2.2.2.9) is determined by the conditions $B \geqslant S_i$, $i = 1, \ldots, m$, in view of which it is nonsingular on the subspace $\mathscr{E}^0$ in the sense that $BD = 0 \Rightarrow D = 0$ for every operator D in $\mathscr{E}^0$. Otherwise, operator $D^+ (B - S_i) D$ could be negative for at least one $i \in 1, \ldots, m$. This last fact means that the first condition in (2.2.2.8) is met

only if $E = D^0$, that is, the optimal decision operators D_i^0 determine the decomposition of unity $E^0 = \sum_{i=1}^{m} D_i^0$, the orthoprojector on subspace $\mathscr{E}^0$, and operator B^0 can be found uniquely by summation of the remaining optimality conditions in (2.2.2.8).

When the subspaces $\mathscr{K}_i = H_i\mathscr{H}$ have a low dimensionality, say, ordinary amplitudes $H_i = (\psi_i \mid$ for which $\mathscr{K}_i = \mathbb{C}$, it has proved expedient to represent the solution to problem (2.2.2.7) via the following

Theorem 2.2.2.2 *The optimal decision operators D^0 determined by conditions* (2.2.2.8) *for $C_i = H_i^+ H_i$, $i = 1, \ldots, m$, have the following form in space $\mathscr{E}^0$*

$$D_i^0 = (L^0)^{-1} H_i^+ \mu_i^0 H_i (L^0)^{-1} \quad i = 1, \ldots, m, \tag{2.2.2.14}$$

where $L^0 = \left(\sum_{i=1}^{m} H_i^+ \mu_i H_i\right)^{1/2} = B^0$ *is the solution to problem* (2.2.2.9), *and the μ_i are trace class positive operators in $\mathscr{K}_i$ defined by the conditions*

$$(1_i - H_i (L^0)^{-1} H_i^+)\, \mu_i^0 = 0, \quad 1_i \geqslant H_i (L^0)^{-1} H_i^+ \tag{2.2.2.15}$$

(1_i *are the identity operators in $\mathscr{K}_i$). If these conditions are met, maximal intensity of graded signals $\varkappa^0 = \sum_{i=1}^{m} \operatorname{Tr} \mu_i^0$ is achieved.*

Proof. Multiplying the remaining equations in (2.2.2.8) from the right by $(L^0)^{1/2}$ and from the left by $(L^0)^{-1/2}$, where $L^0 = B^0$, we can rewrite the optimality conditions in the form

$$(E^0 - F_i^+ F_i)\, M_i^0 = 0, \quad E^0 \geqslant F_i^+ F_i, \; i = 1, \ldots, m, \tag{2.2.2.16}$$

where $F_i = H_i (L^0)^{-1/2}$ and $M_i^0 = (L^0)^{-1/2} D_i^0 (L^0)^{1/2}$. Thus,

$$M_i^0 = F_i^+ F_i M_i^0 = M_i^0 F_i^+ F_i = F_i^+ \mu_i^0 F_i, \tag{2.2.2.17}$$

where $\mu_i^0 = F_i M_i^0 F_i^+$, which leads to (2.2.2.14) if we carry out the inverse transformation. If we substitute (2.2.2.17) into Eq. (2.2.2.16) and multiply the result from the right by F_i^+ and from the left by $(F_i^+)^{-1}$, we arrive at (2.2.2.15) if we allow for the reversibility of the operators $F^+\colon \mathscr{K}_i \to \mathscr{E}$. The inequalities (2.2.2.15) are simply the inequality (2.2.2.16) in the form $F_i F_i^+ \leqslant 1_i$. The operator L^0 is determined by summation $\sum_{i=1}^{m} D_i^0 = E^0$ of the optimal decision operators (2.2.2.14), which yields $L^0 = \sum_{i=1}^{m} H_i^0 \mu_i^0 H_i$, and this determines uniquely the positive operator $B^0 = L^0$.

The proved theorem reduces the solution of the optimal identification problem to finding the operators μ_i^0 that satisfy conditions

(2.2.2.15), which in the case of finitely mixed patterns H_i constitute finite-dimensional algebraic equations and inequalities. For one, for pure patterns $H_i = (\psi_i \,|$, conditions (2.2.2.15) have the scalar form

$$\mu_i^0 = (\psi_i \mid (L^0)^{-1} \psi_i)\, \mu_i^0, \quad 1 \geqslant (\psi_i \mid (L^0)^{-1} \psi_i),$$
$$i = 1, \ldots, m, \tag{2.2.2.18}$$

where $L^0 = (|\,\psi_i)\, \mu_i^0\, (\psi_i \,|\,)^{1/2}$. The numerical positive solutions of the system of algebraic equations (2.2.2.18) determine the one-dimensional decision operators

$$D_i^0 = |\chi_i)(\chi_i|, \quad \chi_i = (L^0)^{-1}\psi_i \sqrt{\mu_i^0} \tag{2.2.2.19}$$

(which are equal to zero for those i's for which $(\psi_i \mid (L^0)^{-1} \psi_i < 1)$ and the quality of the optimal solution, $\varkappa^0 = \sum_{i=1}^{m} \mu_i^0$.

Solution of the pattern identification problem makes it possible to establish the quasioptimal multialternative detection scheme using the maximum criterion of the total degree of contrast (2.2.2.6) as the first approximation in $\sqrt{\varepsilon}$ for decision operators of the form

$$D_i = (F_{0i} + \sqrt{\varepsilon}\, F_{1i})^+ (F_{0i} + \sqrt{\varepsilon}\, F_{1i}) = F_{\varepsilon i}^+ F_{\varepsilon i}. \tag{2.2.2.20}$$

Assuming that $F_{0i} = \sqrt{\mu_i^0} H_i\, (L^0)^{-1}$ and $D_{0i} = F_{0i}^+ F_{0i}$, in the first order in the signal-to-noise ratio $\varepsilon \ll 1$ we obtain the following formula for the degree of contrast of quasioptimal detection $\varkappa_0 = \sum_{i=1}^{m} S_i D_{0i}$

$$\varkappa_0 = \sum_{i=1}^{m} \mathrm{Tr}_{\mathscr{K}_i} \mu_i^0 (1 + \sqrt{\varepsilon}\,(\gamma_i + \gamma_i^+)/2), \tag{2.2.2.21}$$

where $\gamma_i = H_i A_i^+ (L^0)^{-1} H_i^+$, or $\gamma_i = (A_i \psi_i \,|\, (L^0)^{-1} \psi_i)$ when $\mathscr{K}_i = \mathbb{C}$.

2.2.2.4 The Signal Representation

It has proved expedient to represent solution (2.2.2.14) to the problem of optimal identification of wave patterns in the so-called signal space, $\mathscr{K}^m = \bigoplus_{i=1}^{m} \mathscr{K}_i$, which is the direct sum of Hilbert spaces $\mathscr{K}_i = H_i \mathscr{H}$ and which, in the case of ordinary amplitudes $H_i = (\psi_i \,|$, is equal to $\mathbb{C}^m$. Such decomposition is carried out via the partially isometric operator V: $\mathscr{H} \to \mathscr{K}^m$ of the polar expansion $H = \sigma^{1/2} V$, $\sigma = HH^+$, for the operator H: $\varphi \in \mathscr{H} \mapsto [H_i \varphi]_{i=1}^m$ from $\mathscr{H}$ into $\mathscr{K}^m$, which is defined uniquely on the subspace $\mathscr{E}^0$ by the conditions $V^+V = E^0$ and $VV^+ = \varepsilon^0$, where ε^0 is the support of the correlation matrix $\sigma = \sigma\varepsilon^0$. Note that the m-by-m matrix

$\sigma = [\sigma_{ik}]$ consisting of operator components $\sigma_{ik} = H_i H_k^+$, $i, k = 1, \ldots, m$ ($\sigma_{ik} = (\psi_i \mid \psi_k)$ if $H_i = (\psi_i \mid)$), is positive and, in the case of the linear independence of the signals H_i, nonsingular with support $\varepsilon^0 = \bigoplus_{i=1}^{m} 1_i \equiv 1^m$. The components $V_i = 1_i V$, $i = 1, \ldots, m$, of the isometric operator $V: \mathcal{E}^0 \to \mathcal{K}^m$ determined by the diagonal projectors 1_i from $\mathcal{K}^m$ onto $\mathcal{K}_i$ bring about, obviously, the decomposition of the unit element

$$E^0 = V^+ V = \sum_{i=1}^{m} V^+ 1_i V = \sum_{i=1}^{m} V_i^+ V_i \tag{2.2.2.22}$$

of space $\mathcal{E}^0$ and are orthogonal if $\varepsilon^0 = 1^m$:

$$V_i V_k^+ = \varepsilon_{ik}^0 = 1_i \varepsilon^0 1_k = \delta_{ik} 1_k.$$

Representing the operators H_i in the form $H_i = h_i V$, with $h_i = 1_i \sigma^{1/2}$, we can write the necessary and sufficient conditions for the optimality of the separating operators D_i in the following form:

$$(\mathring{\lambda} - \sigma_i)\, \delta_i^0 = 0, \ \mathring{\lambda} \geqslant \sigma_i = h_i^+ h_i, \ i = 1, \ \ldots, \ m, \tag{2.2.2.23}$$

which are simply conditions for the decomposition of the m-by-m projection matrix $\varepsilon^0 = \sum_{i=1}^{m} \delta_i^0$, $\delta_i^0 = VD_i^0 V^+$. Theorem 2.2.2.2 in this case assumes the form of

Theorem 2.2.2.3 *The optimal decomposition of the support ε^0 of the correlation matrix σ defined by conditions* (2.2.2.23) *has the form*

$$\delta_i^0 = \mathring{\lambda}^{-1} h \mu_i^0 h \mathring{\lambda}^{-1}, \ i = 1, \ \ldots, \ m, \tag{2.2.2.24}$$

where $h = \sigma^{1/2}$, $\mathring{\lambda} = (h\mu^0 h)^{1/2}$, and $\mu^0 = \bigoplus_{i=1}^{m} \mu_i$ is a diagonal matrix $\mu^0 = [\mu_i^0 \delta_{ik}]$ consisting of the positive operators $\mu_i^0: \mathcal{K}_i \to \mathcal{K}_i$ and defined by the conditions

$$\mu^0 = \iota\,(h\mathring{\lambda}^{-1}h)\,\mu^0, \ 1^m \geqslant \iota\,(h\mathring{\lambda}^{-1}h), \tag{2.2.2.25}$$

or $\mu^0 = \iota(\sqrt{\sigma\mu^0})$ if σ is nonsingular ($\varepsilon^0 = 1$), *where $\iota: a \mapsto \sum_{i=1}^{m} 1_i a 1_i$ is the diagonalization operation $[a_{ik}] \mapsto [a_{ik}\delta_{ik}]$ of the m-by-m matrices $a = [a_{ik}]$ consisting of operators $a_{ik}: \mathcal{K}_k \to \mathcal{K}_i$. The quality of optimal identification is determined by the trace in $\mathcal{K}^m$, or $\varkappa^0 =$ Tr μ^0.*

Proof. Representation (2.2.2.24) can be obtained directly via the isomorphism V of spaces $E^0 \mathcal{H}$ and $\varepsilon^0 \mathcal{K}^m$. Here $\mathring{\lambda}^{-1} = V\,(L^0)^{-1} V^+$, an m-by-m matrix with elements $(\mathring{\lambda}^{ki})^{-1}: \mathcal{K}_i \to \mathcal{K}_k$, is the inverse

of matrix $\overset{\circ}{\lambda} = VL^0V^+$ with respect to ε^0: $\overset{\circ}{\lambda}\overset{\circ}{\lambda}^{-1} = \varepsilon^0 = \overset{\circ}{\lambda}^{-1}\overset{\circ}{\lambda}$. Matrix $\overset{\circ}{\lambda} = VL^0V^+$ consisting of operator elements $\overset{\circ}{\lambda}_{ik}$: $\mathcal{K}_k \to \mathcal{K}_i$ is directly expressible in terms of the square root $h = \sqrt{\sigma}$ of the correlation matrix

$$\overset{\circ}{\lambda} = \left(\sum_{i=1}^{m} h_i^+ \mu_i^0 h_i\right)^{1/2} = \sqrt{h\mu^0 h}, \tag{2.2.2.26}$$

while the conditions (2.2.2.15) for determining the operators μ_i^0, which in the signal representation have the form

$$(1_i - h_i\overset{\circ}{\lambda}^{-1}h_i^+)\mu_i^0 = 0, \quad 1_i \geqslant h_i\overset{\circ}{\lambda}^{-1}h_i^+, \tag{2.2.2.27}$$

represent the element-by-element notation for the conditions (2.2.2.25) imposed on the diagonal elements in $\mathcal{K}^m$. If σ is nonsingular (which means that h is nonsingular, too), we can rewrite Eq. (2.2.2.25) in the following simple form:

$$\mu^0 = \iota\,(h\overset{\circ}{\lambda}^{-1}h\mu^0) = \iota\,(h\overset{\circ}{\lambda}h^{-1}) = \iota\,(\sqrt{\delta\mu^0}), \tag{2.2.2.28}$$

where we have allowed for the fact that $\iota\,(a)\,\mu^0 = \iota\,(a\mu^0)$ (because μ^0 is diagonal) and that $\sigma\mu^0 = (h\overset{\circ}{\lambda}h^{-1})^2$, in accordance with (2.2.2.26). The proof of Theorem 2.2.2.3 is complete.

We note an important particular case when conditions (2.2.2.25) can be resolved explicitly. Let us assume that the diagonal part $\iota\,(h)$ of matrix $h = \sigma^{1/2}$ is commutative with h. Then conditions (2.2.2.25) are met at $\mu^0 = \iota\,(\sqrt{\sigma})^2$, that is, at $\mu_i^0 = (h_{ii})^2$, $i = 1, \ldots, m$. Indeed, the diagonal matrix μ^0 in this case is commutative with h and

$$\overset{\circ}{\lambda} = \sqrt{h\mu^0 h} = \sqrt{h^2\mu^0} = h\sqrt{\mu^0} = h\,\iota\,(h) = \sqrt{\sigma}\iota\,(\sqrt{\sigma}).$$

Moreover, $h\overset{\circ}{\lambda}^{-1}h = h\iota\,(h)^{-1}$, where $\iota\,(h)^{-1}$ is the diagonal that is the inverse of $\iota\,(h)$, which always exists because the diagonal elements $\sigma_{ii} = H_iH_i^+$ of the correlation matrix σ are nonsingular and, hence, so are the diagonal elements h_{ii} of the matrix $h = \sqrt{\sigma}$ on the spaces $\mathcal{K}_i = H_i\mathcal{H}$. Thus,

$$\iota\,(h\overset{\circ}{\lambda}^{-1}h) = \iota\,(h\iota\,(h)^{-1}) = \iota\,(h)\,\iota\,(h)^{-1} = 1^m,$$

and conditions (2.2.2.25) are satisfied. The optimal decision operators then assume the form

$$\delta_i^0 = 1_i, \ D_i^0 = V^+1_iV, = V_i^+V_i, \ i = 1, \ \ldots, \ m, \tag{2.2.2.29}$$

where $V = h^{-1}H$, while the quality of optimal separation is determined by the total intensity:

$$\varkappa^0 = \sum_{i=1}^{m} \operatorname{Tr}(h_{ii})^2 = \sum_{i=1}^{m} \operatorname{Tr}(\sigma_{ii}^{1/2})^2. \tag{2.2.2.30}$$

The above-noted property of commutativity manifests itself, for one thing, in the case where all diagonal operators σ_{ii} coincide and are multiples of the identity element $1_i = 1$ of space $\mathscr{K}_i = \mathscr{K}$, which is the same for all $i = 1, \ldots, m$. In view of the assumption that σ_{ii} is a trace class operator and, hence, $\mu_i^0 = (\sigma_{ii})^2$, this is possible only for a finite-dimensional $\mathscr{K}$. In Section 2.2.2.5 we consider concrete equidiagonal families of ordinary amplitudes $H_i = (\psi_i|$, for which $\mathscr{K} = \mathbb{C}$.

2.2.2.5 Separation of Cyclic Systems

Let $\{\psi_i\}_{i=1}^m$ be a family (or set) of nonorthogonal wave amplitudes $\psi_i \in \mathscr{H}$ that describe sound or visual patterns with a correlation matrix $\sigma = [(\psi_i \mid \psi_k)]$ whose square root, $h = \sqrt{\sigma}$, has the same diagonal elements $h_{ii} = a = h_{kk}$ for all $i, k = 1, \ldots, m$. The optimal identification of the wave patterns $\{\psi_i\}$ is described by the one-dimensional separating operators $\{D_k^0\}_{k=1}^m$ of the form (2.2.2.12), where $\chi_k^0 = V_k^+$, $k = 1, \ldots, m$, is generally an overcomplete system of polar decomposition,

$$\psi_k = \sum_{i=1}^{m} V_k^+ \sigma_{ki}^{1/2} = \sum_{i=1}^{m} \chi_k^0 h_{ki},$$

$$\sum_{k=1}^{m} |\chi_k^0)(\chi_k^0| = \sum_{k=1}^{m} V_k^+ V_k = V^+ V = E^0,$$

in the space $\mathscr{E}^0$ induced by the set $\{\psi_k\}$. Bearing in mind that $\mu = a^2 = (\operatorname{Tr} h/m)^2$, we can represent the maximal intensity $\varkappa^0 = ma^2$ of optimally separated amplitudes $\{\chi_i^0\}$ in the following invariant form:

$$\varkappa^0 = (\operatorname{Tr} h)^2/m = (\operatorname{Tr} \sigma^{1/2})^2/m.$$

Let us consider the following example when the above-mentioned condition of the equidiagonality of matrix $h = \sqrt{\sigma}$ is met. We will call the system $\{\psi_i\}$ of amplitudes of equal intensity $\|\psi_i\|^2 = \nu$ equiangular if $(\psi_i \mid \psi_k) = \nu\gamma$ for every $i \neq k$, that is, if the cosines of mutual angles are equal to γ. This is possible in the case when $\gamma \geqslant 1/(1-m)$, say, when $\psi_i = \varphi_0 + \varphi_i$, where $\{\varphi_i\}_{i=0}^m$ is an orthogonal system of amplitudes with intensities $\|\varphi_0\|^2 = \nu\gamma$ and $\|\varphi_i\|^2 = \nu(1-\gamma)$ at $i \neq 0$. Representing the respective correlation matrix σ in the form

$$\sigma = \nu((1-\gamma)1^m + \gamma x^{\mathrm{T}} x),\ x = (1, \ldots, 1) \in \mathbb{C}^m, \tag{2.2.2.31}$$

and using the formula

$$f(1^m + \tau x^{\mathrm{T}} x) = f(1)\, 1^m + \frac{1}{xx^{\mathrm{T}}} \left(f(1 + \tau x x^{\mathrm{T}}) - f(1)\right) x^{\mathrm{T}} x$$

to invert it and extract a square root, we can write out the optimal system $\{\chi_i^0\}$ for $\gamma \in \,](1 - m)^{-1}, 1[$ explicitly:

$$\chi_k^0 = \frac{1}{\sqrt{1-\gamma}} \left(\psi_k/\sqrt{\mu} - (1 - (1 + m\gamma/(1-\gamma))^{-1/2}) \frac{1}{m} \sum_{i=1}^{m} \psi_i/\sqrt{\mu} \right).$$

The intensity of the signals separated by this orthogonal system is

$$\varkappa^0 = \nu \left(m - (1 - 1/m)\left(\sqrt{1-\gamma+m\gamma} - \sqrt{1-\gamma}\right)\right)^2,$$

and admits the maximal value $\varkappa^0 = m\nu$ in the event of orthogonality $\gamma = 0$ of the family $\{\psi_i\}$ and the value $\varkappa^0 = \mu$ in the case of colinearity of $\{\psi_k\}$.

For one, when the $\psi_i = |\,\alpha_i)$ are canonical equiangular amplitudes α_i defined by a $(d+1)$-by-$(d+1)$ matrix of the scalar products of vectors $\alpha_i \in \mathbb{C}^{d+1}$ of the form $\alpha_i^* \alpha_k^{\mathrm{T}} = \lambda\delta$ for $i \neq k$, $|\,\alpha_i\,|^2 = \lambda$ for all $i = 1, \ldots, m$, the quantity $\gamma = \exp\{\lambda(\delta - 1)\}$ does not vanish and the maximal intensity of optimal separation is always lower than $m\mu$ even if the vectors $\{\alpha_i\}$ are orthogonal ($\delta = 0$) and tends to $m\gamma$ only as $\lambda \to \infty$. Note that the maximal intensity of separation of canonical amplitudes is reached on simplex vectors $\alpha_i \in \mathbb{C}^{d+1}$ defined by the condition $\delta = (1 - m)^{-1}$; for one thing, at $m = 2$ the intensity of separation of a pair of canonical amplitudes,

$$\varkappa^0 = \nu\left(1 - \sqrt{1-\gamma^2}\right) = 1 - \sqrt{1 - e^{2\lambda(\delta-1)}},$$

can be attained at $\delta = -1$ by employing orthogonal vectors α_i, while at $\delta = 0$ this can be done only by doubling $\lambda = |\,\alpha_i\,|^2$.

Equiangular systems constitute a particular case of cyclic systems, which are defined by the condition that the correlation matrix σ_{ik} remain invariant under translations $s \in \mathbb{Z}$: $(i, k) \mapsto (i + s, k + s)$, that is, at $\sigma_{ik} = \sigma(i - k)$. Such translation invariant systems as containing only a finite number m of distinct amplitudes must satisfy also the cyclicity condition $\sigma(l) = \sigma(l + s)$ for $l = i - k < 0$. Since the matrix $h = \sqrt{\sigma}$, as any other matrix function of σ, also depends solely on the difference in the indices, or $h_{i,k} = h(i - k)$, the equidiagonality condition $h_{i,k} = a = h(0)$ is certain to be met and the solution to the problem of separating any cyclic system can be written explicitly.

Let us take the case of cyclic canonical amplitudes $\psi_k = |\,\alpha_k)$ defined by complex numbers $\alpha_k \in \mathbb{C}$ whose real and imaginary parts can be interpreted as the mean frequency and duration of the wave packet $|\,\alpha_i)$. There can be only two cases of the cyclicity of amplitudes

$|\alpha_i)$ corresponding to the equidistant distribution of points α_i along a circle or a straight line with the center at $\alpha = 0$.

(1) *Optimal estimation of phase.* Let $\alpha_i = \sqrt{\lambda} e^{2\pi ijk/m}$, $j = \sqrt{-1}$. In this case we have a cyclic system

$$\sigma_{ik} = \exp\{\lambda(e^{-2\pi j(i-k)/m} - 1\} = \sigma(i-k).$$

To extract the square root of matrix σ, we must carry out a discrete Fourier transformation via a unitary matrix $U_{in} = \exp\{2\pi ijn/m\}/\sqrt{m}$, $n = 0, \ldots, m-1$. A continuous analog of this problem, to which one can pass if m is sent to infinity, is the estimation of phase θ of the vector $\alpha_\theta = \sqrt{\lambda}\exp\{2\pi j\theta\}$ of the canonical amplitude $\psi_\theta = |\alpha_\theta)$ on the interval $[0, 1]$. Diagonalizing matrix

$$\sigma_{x\theta} = (\alpha_x|\alpha_\theta) = \exp\{\lambda(e^{-2\pi j(x-\theta)} - 1)\}$$

via a discrete-continuous Fourier transformation $u_{xn} = \exp\{2\pi jxn\}$, $n \in \mathbb{Z}$, we obtain its eigenvalues

$$\lambda_n = \lambda^n e^{-\lambda}/n!, \quad n = 0, 1, \ldots, \quad \lambda_n = 0, \quad n < 0.$$

The optimal system of decision vectors χ_x^0, $x \in [0, 1]$, has the form

$$\chi_x^0 = \sum_{n=0}^{\infty} e^{2\pi jxn}|n), \text{ where } |n) = \frac{1}{\sqrt{n!}}(A^+)^n|0),$$

with A^+ the creation operator in $\mathcal{H} = \mathcal{L}^2(\mathbb{R})$. It can easily be verified that the system χ_x^0 defines the decomposition of unity,

$$I = \int_0^1 |\chi_x^0)(\chi_x^0|\,dx = \sum_{n,m=0}^{\infty} |n)(m| \int_0^1 e^{2\pi jx(n-m)}\,dx = \sum_{n=0}^{\infty} |n)(n|,$$

but is not orthogonal.

(2) *Optimal estimation of amplitude.* Let us take $\alpha_i = i\Delta e^{j\theta}$, where $i \in \mathbb{Z}$, $\Delta > 0$, and $j = \sqrt{-1}$. In this case the cyclicity condition is satisfied:

$$\sigma_{ik} = \exp\{-\Delta^2(i-k)^2/2\} = \sigma(i-k).$$

The matrix $\sigma = [\sigma_{ik}]$ is diagonalized by the discrete-continuous Fourier transformation $U_{i\lambda} = \exp\{2\pi ji\lambda\}$, $\lambda \in [0, 1]$. For $\Delta \ll 1$ the problem of optimal separation of the respective coherent amplitudes is reduced to the problem of optimal estimation of the real parameter $x \in \mathbb{R}$ of the coherent amplitude $\psi_x = |xe^{j\theta})$. This estimation is realized by measuring the intensity in the proper representation of the self-adjoint operator $\operatorname{Re} Ae^{-j\theta}$ in space $\mathcal{H} = \mathcal{L}^2(\mathbb{R})$. At $\theta = 0$ this is the frequency representation, while at $\theta = \pi/2$ it is the temporal representation.

2.2.3 Optimal Testing and Discrimination of Mixed Wave Hypotheses

In this section we will take up the problem of testing wave hypotheses based on measuring the appropriate intensity distributions. We will derive the necessary and sufficient conditions for optimal testing of such hypotheses by the minimum criterion of parasitic contrast at a fixed level of the received signal by employing a method of linear programming in partially ordered Banach operator spaces. In a specific case these conditions formally coincide with conditions obtained earlier in [2.15] on the optimality of quantum measurements by the minimum criterion for the error probability. A general geometric solution will be given for the case of a two-dimensional space, which is sufficient for describing the recognition of the polarization of a plane wave. This solution is similar to solution of the problem of measuring quantum mechanical spin [2.4].

2.2.3.1 Wave Hypotheses

The problem of recognizing sound and visual patterns based on measurements of the intensity of the received audio or optical wave can be formulated within the framework of the wave theory of hypothesis testing discussed below.

Let H_i, $i = 1, \ldots, m$, be bounded operators from the Hilbert space $\mathscr{H}$ to another Hilbert space $\mathscr{K}$ describing the possible generalized random amplitudes at the "in" terminals of the receiver with density operators $S_i = H_i^+ H_i$ with a trace $\operatorname{Tr} S_i < \infty$. The reader will recall that at $\mathscr{K} = \mathbb{C}$ the operators H_i correspond to ordinary amplitudes $\psi_i \in \mathscr{H}$, $i = 0, \ldots, m$, which define the bounded functionals $H_i = (\psi_i \,|\colon \chi \in \mathscr{H} \mapsto (\psi_i \mid \chi)$. Each operator H_i, $i = 0, \ldots, m$, can be thought of as a hypothesis, according to which at the "in" terminals of the receiver there is one of the possible simple or mixed patterns G_i, $i = 1, \ldots, m$, in a partially coherent superposition $H_i = G_i + H_0$ with an acoustic or optical background described by operator H_0 in the absence of wave patterns G_i. The problem of m-alternative detection of wave patterns G_i, $i = 1, \ldots, m$, may, therefore, be considered as a problem of testing $m + 1$ hypotheses H_i, $i = 0, \ldots, m$, and vice versa.

The optimal testing of the hypotheses H_i, $i = 0, \ldots, m$, is determined by the solution to the problem of finding a quasiselective measurement $D = \{D_i\}_{i=0}^m$ that maximizes the quality functional

$$\varkappa(R, D) = \sum_{i=0}^{m} \langle R_i, D_i \rangle, \quad D_i \geqslant 0, \quad \sum_{i=0}^{m} D_i = E,$$

where $R = \{R_i\}_{i=0}^m$ are trace class operators with a support E: $R_i E = R_i$ for all $i = 0, \ldots, m$, operators that are usually represented by linear combinations $R_i = \sum_{k=0}^{m} c_i^k S_k$ of density operators $S_i = H_i^+ H_i$. For instance, in the problem of m-alternative optimal

detection by the maximum criterion for the total degree of contrast (2.2.2.7), the operators R_i are in effect the degrees of contrast $R_0 = 0$, $R_i = S_i - S_0 = C_i$, $i = 1, \ldots, m$, and the admissible operators $\{D_i\}_{i=0}^m$ are determined by the decision operators $D_0 = E - D$, D_i, $i = 1, \ldots, m$, where $D = \sum_{i=1}^m D_i$. For the problem of discriminating between the hypotheses H_i we can consider more general criteria defined, say, by the operators

$$R_0 = \sum_{i=1}^m C_i, \quad R_i = (1+\lambda_i) C_i, \quad \lambda_i \geqslant 0, \quad i = 1, \ldots, m, \tag{2.2.3.1}$$

that appear in the problem of suppressing parasitic degrees of contrast $\langle C_i, D_k \rangle$ for $i \neq k$:

$$\tau^0(\mathbf{C}) = \inf_{D_i \geqslant 0} \left\{ \sum_{i=1}^m \left\langle C_i \sum_{k \neq 1} D_k \right\rangle \,\middle|\, \langle C_i, D_i \rangle \geqslant \varepsilon_i, \ \sum_{i=1}^m D_i \leqslant E \right\} \tag{2.2.3.2}$$

under the condition that the useful degrees of contrast $\langle C_i, D_i \rangle$, $i = 1, \ldots, m$, are not lower than given levels ε_i. Indeed, if we solve the extremal problem

$$\begin{aligned}
\varkappa^0(R) &= \sup_{D_i \geqslant 0} \left\{ \sum_{i=0}^m \langle R_i, D_i \rangle \,\middle|\, \sum_{i=1}^m D_i \leqslant E \right\} \\
&= \sup_{D_i \geqslant 0} \left\{ \sum_{i=1}^m (\langle C_i, D_i \rangle (1+\lambda_i) \right. \\
&\qquad \left. + \langle C_i, E - D \rangle) \,\middle|\, \sum_{i=1}^m D_i \leqslant E \right\}
\end{aligned} \tag{2.2.3.3}$$

for the operators $R_i = R_i^\lambda$ defined in (2.2.3.1), we can write the solution to the problem (2.2.3.2) in the form

$$\begin{aligned}
\tau^0(\mathbf{C}) &= \sum_{i=1}^m \langle C_i, E \rangle + \sup_{\lambda_i \geqslant 0} \left\{ \sum_{i=1}^m \lambda_i \varepsilon_i - \varkappa^0(R^\lambda) \right\} \\
&= \sup_{\lambda_i \geqslant 0} \inf_{D_i \geqslant 0} \left\{ \sum_{i=1}^m \left(\left\langle C_i, \sum_{k \neq i} D_k \right\rangle \right. \right. \\
&\qquad \left. \left. + \lambda_i (\varepsilon_i - \langle C_i, D_i \rangle) \right) \,\middle|\, \sum_{i=1}^m D_i \leqslant E \right\},
\end{aligned} \tag{2.2.3.4}$$

provided that we employ Lagrange's method of multipliers λ_i, $i = 1, \ldots, m$.

Let us start with the classical variant

$$\varkappa_M^0(R) = \sup_{\Delta_i \geqslant 0} \left\{ \sum_{i=0}^{m} \langle R_i, M(\Delta_i) \rangle \,\middle|\, \sum_{i=0}^{m} \Delta_i = X \right\}$$

$$= \sum_{i=0}^{m} \mu_i (\Delta_i^0) \qquad (2.2.3.5)$$

of the problem (2.2.3.3) of optimal testing of hypotheses H in a fixed measurement described by the decomposition $E = \int M(dx)$ of an orthoprojector E, $R_i E = R_i$, on a Borel space X. This may be the coordinate selective measurement $M(dx) = I(dx)$, $X = \Omega$, or x the momentum quasiselective measurement $M(dx) = \widetilde{I}(dx)$, $X = \mathbb{R}^{d+1}$, or the canonical quasimeasurement $M(dx) = |x)(x|\,dx$, $X = \mathbb{C}^{d+1}$, described in Section 2.1.2. The upper bound (2.2.3.5) in measurable partitions $X = \sum_{i=0}^{m} \Delta_i$ reaches the gauge

$$\langle \mu \rangle = \inf \{ \lambda(X) \,|\, \lambda \geqslant \mu_i,\ i = 0, \ldots, m \} = \mu_V(X) \quad (2.2.3.6)$$

of the family $\mu = \{\mu_i\}_{i=0}^m$ of measures $\mu_i(\Delta) = \langle C_i, M(\Delta) \rangle$, where the infimum is taken over all the measures of finite variation $|\lambda|(X) < \infty$ that majorize all μ_i.

Indeed, $\varkappa_M^0(C) \leqslant \varkappa_V(\mu)$, since for every measurable partition $\Omega = \sum_{i=0}^{m} \Delta_i \subseteq X$, obviously,

$$\sum_{i=0}^{m} \mu_i(\Delta_i) \leqslant \sum_{i=0}^{m} \lambda(\Delta_i) = \lambda(\Omega) \leqslant \lambda(X).$$

The lower bound (2.2.3.6) is attained at the upper bound $\mu_V = \bigvee_{i=0}^{m} \mu_i$ of the family of measures $\{\mu_i\}_{i=0}^m$: $\mu_V \geqslant \{\mu_i\}_{i=0}^m$, $\lambda \geqslant \mu_i \Rightarrow \lambda \geqslant \mu_V$, and is equal to the supremum (2.2.2.5) reached on the partitions $\Omega = \sum_{i=0}^{m} \Delta_i^0$ of the support $\Omega \subseteq X$ of measure μ_V into regions Δ_i^0, on which it coincides with the respective measure μ_i:

$$\mu_V(\Delta_i^0) = \max_{k=0,\ldots,m} \mu_k(\Delta_i^0) = \mu_i(\Delta_i^0). \qquad (2.2.3.7)$$

In view of the last relationship, determining a hypothesis H_i for a given measurement M is reduced to searching for the number of the nonempty region Δ_i^0 on which the measured degree of contrast

reaches the envelope $\mu_V(\Delta_i^0)$ of the family $\{\mu_i\}$. However, this method does not enable us to find the wave patterns for which $\mu(\Delta_i^0) < \mu_V(\Delta_i^0)$ for all $i = 0, \ldots, m$.

2.2.3.2 Optimal Testing

To obtain a satisfactory solution to the problem of wave pattern recognition one must look for the supremum (2.2.3.5) not only over the measurement regions Δ_i but also over the various methods of such a measurement, which are described by the resolving operators $D_i = M(\Delta_i)$. Thus, there emerges a nonclassical extremal problem (2.2.3.3), which may be considered as part of the conditional problem (2.2.3.2) of testing the hypotheses H_i in the degrees of contrast $\langle C_i, D_i \rangle$, which are compared with given levels ε_i, $i = 1, \ldots, m$. The necessary and sufficient conditions for solving this problem are formulated in the following

Theorem 2.2.3.1 *The upper bound* (2.2.3.3) *is attained on operators* D_i^0, $i = 0, \ldots, m$, *if and only if there is a trace class operator* $L^0 \geqslant R_i$, $i = 0, \ldots, m$, *such that*

$$(L^0 - R_i) D_i^0 = 0, \quad i = 0, \ \ldots, \ m. \tag{2.2.3.8}$$

The operator L^0 *is then the solution to the duality problem*

$$\langle R \rangle_+ = \inf_L \{\langle L, E \rangle \mid L \geqslant R_i, \ i = 0, \ \ldots, \ m\} \tag{2.2.3.9}$$

for which conditions (2.2.3.8) *are also necessary and sufficient for* $D_i^0 \geqslant 0$, $\sum_{i=0}^{m} D_i^0 = E$, *and* $\varkappa^0(R) = \langle R \rangle_+$. *Solution* L^λ *to this problem for operators* $R_i = R_i^\lambda$, $i = 0, \ldots, m$, *of the form* (2.2.3.1) *represents the solution to the conditionally extremal problem* (2.2.3.2) *in the form*

$$\tau^0(\mathbf{C}) = \sum_{i=1}^{m} (\langle C_i, E \rangle + \mathring{\lambda}_i \varepsilon_i) - \langle L^\lambda, E \rangle, \tag{2.2.3.10}$$

and the parameters $\mathring{\lambda} \geqslant 0$ *can be found from*

$$\mathring{\lambda}_i (\varepsilon_i - \langle C_i, D_i^{\mathring{\lambda}} \rangle) = 0, \ \varepsilon_i \leqslant \langle C_i, D_{i\mathring{\lambda}} \rangle, \ i = 1, \ \ldots, \ m, \tag{2.2.3.11}$$

where D_i^λ *are the optimal decision operators at* $R_i = R_i^\lambda$.

Proof. The sufficiency of the optimality conditions (2.2.3.8) for (2.2.3.7) and (2.2.3.9) can be verified directly by employing the property of monotonicity of a trace, $L \geqslant R_i \Rightarrow \mathrm{Tr}(LD_i) \geqslant \mathrm{Tr}(R_i D_i)$, for $D_i \geqslant 0$. Allowing for the equality $L^0 E = \sum_{i=0}^{m} R_i D_i^0$, which is obtained via summation of (2.2.3.8) over $i = 0, \ldots, m$,

for every family $\{D_i\}_{i=0}^{m}$ admissible in (2.2.3.7) we have

$$\sum_{i=0}^{m} \langle R_i,\ D_i\rangle = \sum_{i=0}^{m} \mathrm{Tr}\,(R_i D_i) \leqslant \sum_{i=0}^{m} \mathrm{Tr}\,(L^0 D_i)$$
$$= \mathrm{Tr}\,(L^0 E) = \sum_{i=0}^{m} \langle R_i,\ D_i^0\rangle.$$

In a similar manner for every operator L admissible in (2.2.3.9) we have

$$\langle L,\ E\rangle = \mathrm{Tr}\,(LE) = \sum_{i=0}^{m} \mathrm{Tr}\,(LD_i^0) \geqslant \sum_{i=0}^{m} R_i D_i^0 = \langle L^0,\ E\rangle.$$

The necessity of the optimality conditions (2.2.3.8) follows from Lagrange's duality principle

$$\sup_{D_i \geqslant 0} \left\{ \sum_{i=0}^{m} \langle R_i,\ D_i\rangle \,\middle|\, \sum_{i=0}^{m} D_i = E \right\}$$
$$= \sup_{D_i\geqslant 0}\ \inf_{L} \left\{ \sum_{i=0}^{m} \langle R_i,\ D_i\rangle + \left\langle L,\ E - \sum_{i=0}^{m} D_i \right\rangle \right\}$$
$$= \inf_{L}\ \sup_{D_i\geqslant 0} \left\{ \sum_{i=0}^{m} \langle R_i - L,\ D_i\rangle + \langle L,\ E\rangle \right\}$$
$$= \inf_{L} \{\langle L,\ E\rangle \,|\, L \geqslant R_i,\ i = 0,\ \ldots,\ m\},$$

according to which $\sum_{i=0}^{m} \langle R_i,\ D_i\rangle = \varkappa^0(R) = \langle R\rangle = \langle L^0,\ E\rangle$ and

$$\sum_{i=1}^{m} \mathrm{Tr}\,[(L^0 - R_i)\, D_i^0] = \mathrm{Tr}\,(L^0 E) - \sum_{i=0}^{m} \langle R_i,\ D_i^0\rangle.$$

The necessary and sufficient condition for this sum of traces of products of positive operators to vanish is, obviously, Eq. (2.2.3.8).

Employment of the duality principle in the conditional problem (2.2.3.2) reduces this problem by the elementary Lagrange method to problem (2.2.3.4), for which the necessity and sufficiency of conditions (2.2.3.11) can be verified directly. The proof of the theorem is complete.

Note that the above proof remains unchanged in the case of an infinite number of hypotheses, m ∞. From this theory follows, for one thing, Theorem 2.2.2.1 if we put $R_0 = 0$, $R_i = C_i$, $i = 1, \ldots, m$, and $L = B$. The existence of a solution to problem (2.2.3.7) and the uniqueness on the subspace $\mathscr{E} = E\mathscr{H}$ of the solution to problem (2.2.3.9) can be obtained from the proof in Section 2.2.2.2 of these assertions for problems (2.2.2.7) and (2.2.2.9) to which

(2.2.3.7) and (2.2.3.9) are reduced by the substitutions $C_i = R_i - R_0$ and $B = L - R_0$.

In the case of positive R_i's the problem of testing the hypotheses H_i, $i = 0, \ldots, m$, can be solved as a problem of separating $m + 1$ signals $H_i = R_i^{1/2}$ by applying Theorems 2.2.2.2 and 2.2.2.3. For nonpositive R_i's it has also proved expedient to go over to the signal space $\mathscr{K}^{m+1} = \bigoplus_{i=0}^{m} \mathscr{K}_i$, $\mathscr{K}_i = H_i\mathscr{H}$, $i = 0, \ldots, m$, via a partially isometric operator $V\colon \mathscr{H} \to \mathscr{K}^{m+1}$ of polar decomposition $H = \sigma^{1/2}V$, $\sigma = HH^+$ for the operator $H\colon \varphi \in \mathscr{H} \mapsto [H_i\varphi]_{i=0}^{m}$. As a result, the optimality conditions for the decision operators D_i^0 can be written in the form of conditions imposed on the decomposition $\varepsilon = \sum_{i=0}^{m} \delta_i^0$, $\delta_i^0 = VD_i^0V^+$, of the support $\varepsilon^0 = VV^+$ of the signal correlation matrix $\sigma_{ik} = H_iH_k^+$, $i, k = 0, \ldots, m$:

$$(\mathring{\lambda} - \rho_i)\,\delta_i^0 = 0, \quad \mathring{\lambda} \geqslant \rho_i = \sum_{k=0}^{m} h_k^+ c_i^k h_k, \quad i = 0, \ \ldots, \ m.$$

Here $\mathring{\lambda} = VL^0V^+$, $h_i = 1_i h$, $h = \sqrt{\sigma}$, and c_i^k is the quality matrix, which defines the operators $R_i = \sum_{k=0}^{m} H_k^+ c_i^k H_k$ and which, for a fixed m, it has proved expedient to consider as being a diagonal operator $c_i = \bigoplus_{k=0}^{m} c_i^k 1_k$ in space $\mathscr{K}^{m+1}$ because then the signal matrices $\rho_i = VR_iV^+$ can be represented in the form $\rho_i = hc_ih$.

Even if the amplitudes H_i are ordinary, that is, $H_i = (\psi_i|$, and hence the correlation matrix is a number matrix $\sigma_{ik} = (\psi_i \mid \psi_k)$, it is difficult to write conditions of optimality explicitly for $m > 1$ for a nonsingular matrix σ. Below we will study this problem for the case where the rank of matrix is equal to 2 and, hence, all the operators R_i, L, and D_i can be represented by 2-by-2 matrices in space $\mathscr{E}^1 = \mathbb{C}^2$.

2.2.3.3 Two-dimensional Recognition

To the operators $\{R_i\}$ in the optimization problem (2.2.3.8) we assign Hermitian matrices that can be considered nonnegative without loss of generality. Any 2-by-2 matrix can be decomposed in Pauli matrices, which are

$$\iota = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

$$R = \begin{bmatrix} v + z & x - jy \\ x + jy & v - z \end{bmatrix} = v + x\sigma_x + y\sigma_y + z\sigma_z = v + \hat{\mathbf{r}}, \quad j = \sqrt{-1},$$

where ν, x, y, and z are real if matrix R is Hermitian, and $\hat{\mathbf{r}} = x\sigma_x + y\sigma_y + z\sigma_z$ is a vector operator represented by vector $\mathbf{r} = (x, y, z)$ of three-dimensional real space $\mathbb{R}^3$. The product of $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$, with $\mathbf{r} \in \mathbb{R}^3$ and $\mathbf{s} \in \mathbb{R}^3$, is equal to $\hat{\mathbf{r}}\hat{\mathbf{s}} = \mathbf{r}\cdot\mathbf{s} + j\,(\mathbf{r} \times \mathbf{s})$, where $\mathbf{r}\cdot\mathbf{s}$ and $\mathbf{r} \times \mathbf{s}$ are the scalar and vector products of $\mathbf{r}$ and $\mathbf{s}$. Note that $\mathrm{Tr}\, R = 2\nu$ and $\mathrm{Det}\, R = \nu^2 - |\,\mathbf{r}\,|^2$ (with $|\,\mathbf{r}\,| = \sqrt{\mathbf{r}\cdot\mathbf{r}}$), and that the nonnegativity condition $R \geqslant 0$ assumes the form $\nu \geqslant |\,\mathbf{r}\,|$. Obviously, rank $R = 2$ for $\nu > |\,\mathbf{r}\,|$, rank $R = 1$ at $\nu = |\,\mathbf{r}\,| \neq 0$, and rank $R = 0$ at $\nu = 0$.

The operators $R_i = (\nu_i + \hat{\mathbf{r}}_i)/2$, $i = 0, \ldots, m$, where $|\nu_i| \geqslant |\mathbf{r}_i|$ and $\sum\limits_{i=0}^{m} \nu_i = 1$, can be interpreted as density operators related to the tested wave hypotheses with a priori intensities $\nu_i = \mathrm{Tr}\, R_i$ and represented by vectors $\mathbf{r}_i \in \mathbb{R}^3$, which are known as polarization vectors. A similar problem arises when we must identify the photon polarization or the electron spin [2.4]. Let us assume that polarizations $\{\mathbf{r}_i\}$ satisfy the inequalities

$$|\,\mathbf{r}_k - \mathbf{r}_i\,| > |\,\nu_k - \nu_i\,| \qquad (2.2.3.12)$$

for all $k \neq i$; in the opposite case, that is, $|\,\nu_k - \nu_i\,| \geqslant |\,\mathbf{r}_k - \mathbf{r}_i\,|$, the kth hypothesis dominates the ith hypothesis or vice versa: $R_k > R_i$ or $R_k = R_i$ or $R_k < R_i$, and one of the hypotheses (with the smaller ν) can be ignored.

The decision operators D_i in the Pauli representation $D_i = \delta_i + \hat{\mathbf{d}}_i$ are described by nonnegative numbers $\delta_i \geqslant 0$ and vectors $\mathbf{d}_i \in \mathbb{R}^3$ ($|\,\mathbf{d}_i\,| \leqslant \delta_i$), with the decomposition of unity $\sum\limits_{i=0}^{m} D_i = 1$ assuming the form

$$\sum_{i=0}^{m} \delta_i = 1, \quad \sum_{i=0}^{m} \mathbf{d}_i = 0.$$

The solution of the problem of optimal recognition of polarizations $\mathbf{r}_i$ with intensities ν_i can be reduced to finding a real number $\mathring{\lambda}$ and a vector $\mathbf{l}^0 \in \mathbb{R}^3$ that defines the operator $L^0 = (\mathring{\lambda} + \hat{\mathbf{l}}^0)/2$ that satisfies conditions (2.2.3.8) for a collection of $D_i^0 = \delta_i^0 + \hat{\mathbf{l}}_i^0$, $i = 0, \ldots, m$.

Theorem 2.2.3.2 *The solution to the problem of optimal recognition of polarizations $\{\mathbf{r}_i\}$ satisfying together with $\{\nu_i\}$ condition* (2.2.3.12) *can be found if and only if there is a collection of numbers $\mu_i^0 \geqslant 0$, $i = 0, \ldots, m$, such that*

$$\left|\sum_{k=0}^{m} (\mathbf{r}_i - \mathbf{r}_k)\,\mu_k^0\right| + \sum_{k=0}^{m} (\nu_i - \nu_k)\,\mu_k^0 \geqslant 1, \quad i = 0, \ldots, m, \qquad (2.2.3.13)$$

where the equality takes place at least for those i's for which $\mu_i^0 \neq 0$. *The optimal decision operators have the form* $\delta_i^0 = |\mathbf{d}_i^0|$, $\mathbf{d}_i^0 = \mu_i^0 (\mathbf{r}_i - \mathbf{l}_i^0)$, *where*

$$\mathbf{l}^0 = \sum_{i=0}^{m} \mu_i^0 \mathbf{r}_i \Big/ \sum_{i=0}^{m} \mu_i^0, \tag{2.2.3.14}$$

and the maximal received intensity is

$$\varkappa^0 = \Big(1 + \sum_{i=0}^{m} \mu_i^0 \nu_i\Big) \Big/ \sum_{i=0}^{m} \mu_i^0 = \mathring{\lambda}. \tag{2.2.3.15}$$

Proof. In terms of $\mathring{\lambda}$, $\mathbf{l}^0$, ν_i, and $\mathbf{r}_i$, the first optimality condition (2.2.3.8), $L^0 - R_i \geqslant 0$, has the form

$$\hat{\lambda} \geqslant |\mathbf{r}_i - \mathbf{l}^0| + \nu_i. \tag{2.2.3.16}$$

The equation $(L^0 - R_i) D_i^0 = 0$ can be written in another form if we nullify the scalar and real vector parts of the product $(\mathring{\lambda} + \nu_i - \hat{\mathbf{l}}^0 - \hat{\mathbf{r}}_i)(\delta_i^0 + \hat{\mathbf{d}}_i)$:

$$(\lambda - \nu_i)\,\mathbf{d}_i^0 + (\mathbf{l}^0 - \mathbf{r}_i)\delta_i^0 = 0, \quad (\mathring{\lambda} - \nu_i)\,\delta_i^0 + (\mathbf{l}^0 - \mathbf{r}_i)\cdot \mathbf{d}_i^0 = 0. \tag{2.2.3.17}$$

The imaginary vector equation $j(\mathbf{l}^0 - \mathbf{r}_i) \times \mathbf{d}_i^0 = 0$, $j = \sqrt{-1}$, follows from the real vector equation in (2.2.3.17), according to which either $\mathbf{d}_i^0$ is collinear with $\mathbf{r}_i - \mathbf{l}^0$ for every $\delta_i^0 \geqslant |\mathbf{d}_i^0|$ or $\mathbf{d}_i^0 = 0$. Since $\mathring{\lambda} - \nu_i \neq 0$, Eqs. (2.2.3.17) are equivalent to

$$\mathbf{d}_i^0 = \delta_i^0 (\mathbf{r}_i - \mathbf{l}^0)/(\mathring{\lambda} - \nu_i), \quad ((\mathring{\lambda} - \nu_i)^2 - |\mathbf{r}_i - \mathbf{l}^0|)^2 \delta_i^0 = 0, \tag{2.2.3.18}$$

in the opposite case ($\mathring{\lambda} = \nu_i$ for a certain i in (2.2.3.16)) for $i = k$ we obtain $\mathbf{l}^0 = \mathbf{r}_i$, with the result that inequalities (2.2.3.18) and (2.2.3.12) become incompatible. The optimal decision vector can be written in the form (2.2.3.14), where $\mu_i^0 = \delta_i^0 (\mathring{\lambda} - \nu_i)$ is nonnegative in accordance with the inequalities $\delta_i^0 \geqslant 0$, $\lambda > \nu_i$, and (2.2.3.16), while $\mathbf{l}^0$ is determined by the set $\{\mu_i^0\}$ in accordance with the fact that $\sum_{i=0}^{m} \mathbf{d}_i^0 = 0$. The second equation in (2.2.3.18) implies that inequalities (2.2.3.16) become amplitudes for the values of i for which $\delta_i^0 = \mu_i(\lambda - \nu_i) \neq 0$. Multiplying (2.2.3.16) by $\sum_{i=0}^{m} \mu_i^0$ and finding λ from the condition that $\sum_{i=0}^{m} \delta_i^0 = 1$ for $\delta_i^0 = \mu_i^0(\lambda - \nu_i) = \mu_i^0|\mathbf{r}_i - \mathbf{l}^0| = |\mathbf{d}_i^0| : \lambda = \Big(1 + \sum_{i=0}^{m} \mu_i^0 \nu_i\Big) \Big/ \sum_{i=0}^{m} \mu_i^0$, we get condition (2.2.3.13) for

determining $\{\mu_i^0\}$. Since Tr $L^0 = \overset{\circ}{\lambda}$, the maximal intensiy of (2.2.3.9) is equal to (2.2.3.15). The proof of the theorem is complete.

Note that the equalities in (2.2.3.13) must be true for at least two indices i and k, since there is no such set $\{\mu_i\}$, $\mu_i \neq 0$, for only one subscript i that satisfies the ith inequality. For every pair $\mathbf{r}_i$ and $\mathbf{r}_k$ for which (2.2.3.12) is valid there is a unique solution of the ith and kth equalities in (2.2.3.13) with $\mu_j = 0$ for all $j \neq i, k$, $\mu_i > 0$, $\mu_k > 0$:

$$\mu_i = (|\mathbf{r}_k - \mathbf{r}_i| + \nu_k - \nu_i)^{-1}, \quad \mu_k = (|\mathbf{r}_i - \mathbf{r}_k| + \nu_i - \nu_k)^{-1},$$

but such a set $\{\mu_i\}$ may not satisfy the other inequalities in (2.2.3.13) for $j \neq i, k$. If there exists a pair $\mathbf{r}_i$, $\mathbf{r}_k$ for which all the inequalities in (2.2.3.13) are valid at $\mu_j \neq 0$ only when $j = i, k$, then the optimal decision vectors $\mathbf{d}_j^0$ are zero for $j \neq i, k$ (see (2.2.3.14)) and

$$\mathbf{d}_i^0 = (\mathbf{r}_i - \mathbf{r}_k)/2\,|\mathbf{r}_i - \mathbf{r}_k|, \quad \mathbf{d}_k^0 = (\mathbf{r}_k - \mathbf{r}_i)/2\,|\mathbf{r}_k - \mathbf{r}_i|.$$

Here the optimal decision operators $D_i^0 = |\mathbf{d}_i^0| + \hat{\mathbf{d}}_i^0$ are orthogonal and correspond to an error intensity $\varkappa^0 = (\nu_i + \nu_k)/2 + |\mathbf{r}_i - \mathbf{r}_k|/2$. In the case where the optimal operators D_i^0 are nonzero for more than two i's, they define a nonorthogonal decomposition of unity in the two-dimensional space $\mathscr{E} = \mathbb{C}^2$. We will not try to find a general analytical solution to the system of equation (2.2.3.13) with $\mu_i^0 \neq 0$ for more than two i's; rather, we will give a geometric interpretation of such a solution.

2.2.3.4 Geometric Interpretation

Let us represent the Hermitian operators (2.2.3.10) by points $r = (\nu, x, y, z) = (\nu, \mathbf{r})$ in the four-dimensional Minkowski space $\mathfrak{R}^{1+3}$. To every nonnegative operator there corresponds a point inside the light cone $\nu = |\mathbf{r}|$. In these terms a priori neither the kth nor the ith hypothesis is dominant at $R_i = (\nu_i + \hat{\mathbf{r}}_i)/2$ and $R_k = (\nu_k + \hat{\mathbf{r}}_k)/2$ if and only if the interval $r_i - r_k = (\nu_i - \nu_k, \mathbf{r}_i - \mathbf{r}_k)$ is spacelike. In accordance with (2.2.3.16), point $\mathbf{l}^0 = (\overset{\circ}{\lambda}, \mathbf{l}^0)$, which represents the operator $L^0 = (\lambda^0 + \hat{\mathbf{l}}^0)$, is the apex of the four-dimensional cone

$$\mathscr{C}(l) = \{r = (\nu, \mathbf{r}):\ \nu - \overset{\circ}{\lambda} + |\mathbf{r} - \mathbf{l}^0| = 0\} \qquad (2.2.3.19)$$

covering all the points $r_i = (\nu_i, \mathbf{r}_i)$ and containing the subset $\{r_{j_\alpha}\} \subset \{r_i\}$ of the boundary points r_{j_α} satisfying (2.2.3.16). On the other hand, the optimal points l are only those whose projections $\mathbf{l}$ belong to the convex hull of the boundary subset of the spatial

projections $\mathbf{r}_{j_\alpha}$, $\alpha = 0, \ldots, s$, $s \leqslant m$:

$$\sum_{\alpha=0}^{s} \mathbf{r}_{j_\alpha}\pi_{j_\alpha} = \mathbf{l}, \quad \sum_{\alpha=0}^{s} \pi_{j_\alpha} = 1, \tag{2.2.3.20}$$

where, in accordance with (2.2.3.14), $\pi_{j_\alpha} = \mu_{i_\alpha} \Big/ \sum_{\alpha=0}^{s} \mu_{j_\alpha} \geqslant 0$ ($\mu_i = 0$ if r_i is covered by cone (2.2.3.19): $\nu_i + |\mathbf{r}_i - \mathbf{l}| \leqslant \lambda$). We will say that the subset $\{r_{j_\alpha}\}$ has an apex if the points r_{i_α} lie on the cone: $r_{j_\alpha} \subset \mathscr{C}(l)$ with an apex l whose spatial projection $\mathbf{l}$ is a point on the convex hull $\{\mathbf{r}_{j_\alpha}\}$. In these terms Theorem 2.2.3.2 can be formulated as follows:

Theorem 2.2.3.3 *To solve the problem of optimal recognition of points* $r_i = (\nu_i, \mathbf{r}_i)$, $i = 0, \ldots, m$, *separated by spacelike intervals* (2.2.3.12), *it is necessary and sufficient to find a subset* $\{r_{j_\alpha}\}$ *with an apex* l^0 *belonging to a cone that covers all other points of set* $\{r_i\}$, *that is, to specify a subset of vectors* $\mathbf{r}_{j_\alpha} \subset \{\mathbf{r}_i\}$, $\alpha = 0, \ldots, s$ *whose convex hull contains vector* $\mathbf{l}^0$ *with respect to which the sum* $|\mathbf{r}_{j_\alpha} - \mathbf{l}^0| + \nu_{j_\alpha}$ *is the constant* $\mathring{\lambda}$:

$$|\mathbf{r}_{j_\alpha} - \mathbf{l}^0| + \nu_{j_\alpha} = \mathring{\lambda}, \quad \alpha = 0, \ldots, s, \tag{2.2.3.21}$$

while $|\mathbf{r}_i - \mathbf{l}^0| + \nu_i \leqslant \mathring{\lambda}$ *for all other indices* $i \in \{j_\alpha\}$. *The optimal decision operators are represented by points on the cone* $d_i^0 = (\delta_i^0, \mathbf{d}_i^0)$, $\delta_i^0 = |\mathbf{d}_i^0|$, *with spatial vectors*

$$\mathbf{d}_i^0 = \pi_i^0 (\mathbf{r}_i - \mathbf{l}^0) \Big/ \sum_{i=0}^{m} \pi_i \mathbf{r}_i, \tag{2.2.3.22}$$

where $\pi_i^0 = 0$ *for* $i \notin \{j_\alpha\}$, *and* $\{\pi_{j_\alpha}^0, \alpha = 0, \ldots, s\}$ *is any nonnegative solution to the system of equations* (2.2.3.20). *The minimal intensity in this case is*

$$\varkappa^0 = \sum_{i=0}^{m} (\nu_i + |\mathbf{r}_i - \mathbf{l}^0|)\, \pi_i. \tag{2.2.3.23}$$

Note that every pair of points r_i, r_k separated by a spacelike interval defines, via two equations from (2.2.3.21), $j_\alpha = i, k$, a set of points $\mathbf{l} \in \mathbb{R}^{1+3}$ whose difference of distances to the points $\mathbf{r}_i$ and $\mathbf{r}_k$ is constant:

$$|\mathbf{l}^0 - \mathbf{r}_k| - |\mathbf{l}^0 - \mathbf{r}_i| = \nu_i - \nu_k. \tag{2.2.3.24}$$

These points lie on one of the two sheets of the hyperboloid of revolution with foci at $\mathbf{r}_i$ and $\mathbf{r}_k$ and eccentricity $\varepsilon = |\mathbf{r}_i - \mathbf{r}_k| / |\nu_i - \nu_k| > 1$. Here, if $\nu_i = \nu_k$, the hyperboloid (2.2.3.24) becomes a plane normal to the segment $\mathbf{r}_i\pi_i + \mathbf{r}_k\pi_k$ ($\pi > 0$, $\pi_i + \pi_k = 1$)

at point $(\mathbf{r}_i + \mathbf{r}_h)/2$, while if $\nu_i \neq \nu_h$, we select the sheet in whose plane lies the focus with the higher intensity, ν_i or ν_k. Obviously, if the subset $\{r_{j_\alpha}\}$ has an apex l, the spatial projection $\mathbf{l}$ is the common point of all the hyperboloids (2.2.3.24) corresponding to all the pairs of the set $\{r_{j_\alpha}\}$ that belong to the convex hull $\{\mathbf{r}_{j_\alpha}\}$. We will call this point $\mathbf{l}$ the center of $\{\mathbf{r}_{j_\alpha}\}$. The uniqueness of operator L^0 implies that the apex of set $\{\mathbf{r}_{j_\alpha}\}$ representing L^0 is unique, which means that the center of $\{\mathbf{r}_{j_\alpha}\}$ is unique, too. It can easily be shown that for every vector $\mathbf{l}$ of the convex hull $\{\mathbf{r}_{j_\alpha}\}$ the system of linear equations (2.2.3.20) has a unique solution $\{\pi^0_{j_\alpha}\}$ if and only if vectors $\mathbf{r}_{j_\alpha} - \mathbf{r}_{j_0}$, $\alpha = 1, \ldots, s$, are linearly independent.

2.2.3.5 Simplex Solutions

The reader will recall that a convex hull of a set $\{\mathbf{r}_{j_\alpha}\}$ of points $\mathbf{r}_{j_\alpha}$, $\alpha = 0, 1, 2, 3$, is called an s-simplex (a segment if $s = 1$, a triangle if $s = 2$, a tetrahedron if $s = 3$, and so on) if the vectors $\mathbf{r}_{j_0}$, $\mathbf{r}_{j_\alpha}$, $\alpha = 1, \ldots, s$, are linearly independent. It is well-known that each s-dimensional face (an s-face) of an n-simplex $(n \geqslant s)$ is a simplex, too. We will call a subset that generates a simplex convex hull a simplex subset.

Theorem 2.2.3.4 *The problem of optimal recognition of polarizations $\{\mathbf{r}_i,\ i = 0, \ldots, m\}$ always has a solution that can be described by the simplex set $\{\mathbf{d}^0_{i_\alpha},\ \alpha = 0, \ldots, s\}$, $s \leqslant m$, of the nonzero vectors (2.2.3.22) corresponding to the simplex subset $\{\mathbf{r}_{j_\alpha}\} \subseteq \{\mathbf{r}_i\}$ with a center at $\mathbf{l}^0$ and a maximal sum*

$$\nu_{j_\alpha} + |\mathbf{r}_{j_\alpha} - \mathbf{l}^0| = \max_{i=0,\ldots,m} \{\nu_i + |\mathbf{r}_i - \mathbf{l}^0|\}.$$

This solution is unique if and only if the s-simplex generated by subset $\{\mathbf{r}_{j_\alpha}\}$ is an s-face of the convex hull of all vectors $\mathbf{r}_{j_0}, \ldots, \mathbf{r}_{j_m}$ with a common center $\mathbf{l}^0$.

Proof. By Theorem 2.2.3.3, the solution to the problem considered here is reduced to finding the cone (2.2.3.19) that covers all points $\{\mathbf{r}_i\}$ and has an apex $\mathbf{l}^0$ with a projection lying inside the convex hull of projections $\{\mathbf{r}_{j_\alpha}\}$ of the tangency points r_{j_α}. Obviously, there is always such a cone. Let $n \leqslant m$ be the number of tangency points r_{j_α}, $\alpha = 0, \ldots, n$. If the subset $\{\mathbf{r}_{j_\alpha},\ \alpha = 0, \ldots, s\}$, $s = m$, is a simplex set, the validity of Theorem 2.2.3.4 is obvious. If this subset is not a simplex, then the convex hull $\{\mathbf{r}_{j_\alpha}\}$ can be partitioned into several simplexes with a common vertex $\mathbf{r}_{j_0}$ via diagonal planes $(\mathbf{r}_{j_0}, \mathbf{r}_{j_\alpha}, \mathbf{r}_{j_\beta})$ or diagonal lines $(\mathbf{r}_{j_0}, \mathbf{r}_{j_\alpha})$ when all the vectors $\mathbf{r}_{j_\alpha}$ are coplanar. Hence, the center $\mathbf{l}^0$ is an interior point of one of the

s-simplexes ($s < n \leqslant m$) with apexes $\mathbf{r}_{j_\alpha}$, $\alpha = 0, \ldots, s$, which are the projections of the tangency points and determine the unique positive solution $\{\pi^0_{j_\alpha}\}$ of system (2.2.3.20). The set $\{\mathbf{d}^0_{j_\alpha}\}$ of nonzero vectors (2.2.3.22) is a simplex set if and only if the set $\{\mathbf{r}_{j_\alpha}\}$ is a simplex and determines the optimal solution with maximal quality (2.2.3.21) and minimal error intensity (2.2.3.23). When center $\mathbf{l}^0$ is an interior point of a nonsimplex convex hull of the projections of tangency points, the optimal simplex solution is not unique (the partition into simplexes is not unique) and there are also optimal nonsimplex solutions. But if point $\mathbf{l}^0$ is a boundary point of the convex hull, that is, an interior point of an s-face, the optimal solution is unique if the face is an s-simplex.

Corollary *To solve the problem of optimal testing of several hypotheses in the two-dimensional space $\mathscr{E} = \mathbb{C}^2$, it is sufficient to limit oneself to $s + 1 \leqslant 4$ solutions $j_0, \ldots, j_s$ corresponding to a simplex subset of hypotheses $\mathbf{r}_{j_0}, \ldots, \mathbf{r}_{j_s}$. Each such solution procedure can be realized in an indirect measurement described by an orthogonal decomposition in the observation space $\mathscr{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$.*

Indeed, in the three-dimensional space $\mathbb{R}^3$ there is not a single simplex subset $\mathbf{r}_{j_0}, \ldots, \mathbf{r}_{j_s}$ for $s > 3$ and, therefore, for every m there always exists an optimal decomposition in the two-dimensional space $\mathscr{E}$ consisting of $s - 1 \leqslant 4$ nonzero decision operators $D^0_{j_\alpha} = \delta^0_{j_\alpha} + \hat{\mathbf{d}}^0_{j_\alpha}$ of rank 1. It is well known that each nonorthogonal decomposition of unity in operators $D_{j_0}, \ldots, D_{j_s}$ of rank 1 can be extended to an orthogonal decomposition in an $(s - 1)$-dimensional space $\mathscr{E} \subset \mathscr{H}$. Hence, we can limit ourselves to the four-dimensional measurement space $\mathscr{H}$, which can always be represented as the tensor product of two-dimensional spaces $\mathscr{E}$: $\mathscr{H} = \mathscr{E} \otimes \mathscr{E}$, corresponding to the composition of two identical systems.

Note that the optimal solution may be degenerate (in the sense that a hypothesis $\mathbf{r}_i$ may correspond to $D_i = 0$) even if the set $\mathbf{r}_0, \ldots, \mathbf{r}_m$ is a simplex set ($m \leqslant 3$), for example, at $m = 2$, $\nu_0 = \nu_1 = \nu_2$, if the vectors $\mathbf{r}_0$, $\mathbf{r}_1$ and $\mathbf{r}_2$ form an obtuse triangle.

In conclusion of this section we will consider two particular cases.

(1) *Optimal recognition of pure polarization.* Here the polarizations are normalized to a priori intensities: $|\mathbf{r}_i| = \nu_i$, with the representative points $r_i = (\nu_i, \mathbf{r}_i)$ belonging to the cone $\nu = |\mathbf{r}|$. Expression (2.2.3.21), which determines the subset of points $\mathbf{r}_{j_\alpha}$ of tangency of this cone and the covering cone (2.2.3.19), has the form

$$|\mathbf{r}_{j_\alpha} - \mathbf{l}^0| + |\mathbf{r}_{j_\alpha}| = \mathring{\lambda}. \tag{2.2.3.25}$$

In relation to $\mathbf{r}_{j_\alpha}$, this is the equation of an ellipsoid of revolution with foci at 0 and $\mathbf{l}^0$ and eccentricity $e = |\mathbf{l}^0| / \mathring{\lambda} < 1$. In accordance

with (2.2.3.16), all the other points $\mathbf{r}_i \notin \{\mathbf{r}_{j_\alpha}\}$ must lie inside the ellipsoid. Hence, the problem of optimal recognition of pure polarization is reduced to finding the ellipsoid described about points $\{\mathbf{r}_i\}$ with foci at 0 and $\mathbf{l}^0$, where $\mathbf{l}^0$ is an interior point of the convex hull of the points of tangency $\{\mathbf{r}_{j_\alpha}\}$. The quality $\varkappa^0$ of the optimal solution is equal to the length of the major axis of the ellipsoid, $\mathring{\lambda}$.

(2) *Optimal recognition of equi-intensity polarizations.* The a priori intensities $\nu_i = \nu_0$, $i = 1, \ldots, m$, and the corresponding points are points of the hyperplane $\nu = \nu_0$. The density operators $R_i = (\nu + \mathbf{r}_i)/2$, all having the same trace ν_0, are represented by normalized vectors, $|\mathbf{r}_i| \leqslant \nu_0$. The intersection of the covering cone (2.2.3.19) and the hyperplane $\nu = \nu_0$ is a sphere $|\mathbf{r} - \mathbf{l}^0| = \rho$ of radius $\rho = \mathring{\lambda} - \nu_0$. Hence, the problem of optimal recognition of equiprobable polarizations is reduced to finding a sphere described about all points $\mathbf{r}_i$: $|\mathbf{r} - \mathbf{l}^0| \leqslant \rho$ with radius ρ and centered at $\mathbf{l}^0$, the center belonging to the convex hull of the tangency points $\mathbf{r}_{j_\alpha}$: $|\mathbf{r}_{j_\alpha} - \mathbf{l}^0| = \rho$. The radius $\rho = \mathring{\lambda} - \nu_0$ determines the maximal intensity (2.2.3.15):

$$\varkappa^0 = \rho + \nu_0 = \mathring{\lambda} \tag{2.2.3.26}$$

($\rho \leqslant \nu_0$ since $|\mathbf{r}_i| \leqslant \nu_0$ for all i's). The minimum of intensity (2.2.3.26) is obtained at $\rho = \nu_0$: $\varkappa^0 = 2\nu_0$. This corresponds to the typical equiprobable case $|\mathbf{r}_i| = 1$, when there is at least one simplex subset $\{\mathbf{r}_{j_\alpha}\}$ for which the center $\mathbf{l}^0 = 0$ is an interior point of the simplex.

2.3 Effective Measurement and Estimation of Parameters of Acoustic Signals and Optical Fields

In this section we develop the noncommutative theory of effective measurements and optimal estimation of unknown parameters of wave patterns as applied to problems of sound and visual pattern recognition. We consider two variants of the lower bounds for the variance of the measured parameters, the variants being based on noncommutative generalizations [2.14, 2.31, 2.49] of the Rao-Cramér inequality [2.50], and introduce the notion of canonical states, for which we derive generalized uncertainty relations similar to the quantum mechanical uncertainty relations [2.11, 2.12, 2.31]. We then establish the necessary and sufficient conditions for effective measurements, conditions that extend the conditions of effectiveness of quantum mechanical measurements obtained in [2.31] to the case of classical wave signals and fields. We formulate the necessary and sufficient conditions for the optimality of generalized measurements

of wave patterns, conditions that generalize the respective conditions substantiated for quantum systems in [2.32]. Finally, we investigate the structure of optimal covariant measurements for symmetric wave patterns, which in the case of quantum symmetric fields has been studied in [2.12, 2.29]. The exposition is based largely on the works of Belavkin [2.29-2.31].

2.3.1 Invariant Bounds for the Variance of Parameters of Wave Patterns

We will consider two variants of the lower bound for the variance of parameters of wave patterns, both based on noncommutative generalizations [2.14, 2.31, 2.49] of the Rao-Cramér inequality. In contrast to [2.14, 2.49], these bounds are represented in a form invariant under diffeomorphisms, the form will be used to obtain generalized uncertainty relations and effective measurements of canonical parameters.

2.3.1.1 Classical Bound

In Section 2.2 we considered the problem of recognizing pure or mixed wave patterns taken from a given finite or denumerable standard family. Generally, sound and visual patterns may contain unknown parameters that run through an infinite set of continuous values $\theta \in \Theta$ of finite or denumerable dimensionality. For example, we may not know the mean frequency and the moment when the sound signal appears or the mean position and the wave number of the visual pattern, or we may a priori have no information on the expected amplitudes of the oscillations in the given finite or denumerable family of standard modes.

It is natural to estimate the unknown parameters by the intensity distributions in the representations in which the wave packets with distinct values of θ are clearly separated; for example, the frequency and position can be calculated as the mean values on the coordinate representation, while the mean time of arrival of a signal and the wave number of a wave packet can be calculated in the momentum representation (but not vice versa).

We will call a family of wave packets described in a Hilbert space $\mathscr{H}$ by amplitudes $\{\psi_\theta\}$ resolvable in a representation defined by the decomposition of unity $I = \int M(dx)$ on a given Borel space X if there exists a measurable map $\hat{\theta}: X \to \Theta$ satisfying the condition

$$\int \hat{\theta}(x)\, \mu_\theta(dx) = \theta \int \mu_\theta(dx) \quad \forall\, \theta \in \Theta, \tag{2.3.1.1}$$

where $\mu_\theta(dx) = (\psi_\theta \mid M(dx)\, \psi_\theta)$ is the respective intensity distribution on X. Thus, the resolvability of the family $\{\psi_\theta\}$ means that it is possible to calculate the unknown $\theta \in \Theta$ in a given representation

given the observed distribution of the μ_θ as the mean values of a function $\hat{\theta}(x)$ known as the unbiased estimator of parameters θ.

It is natural to define the quality of the resolvability of family $\{\psi_\theta\}$ by the size of the variance of the unbiased estimator $\hat{\theta}$, assuming that the quality for a given θ is all the higher the smaller the standard deviation from θ in the distribution induced on Θ by the measure μ_θ of the wave packet ψ_θ. To find the lower bound for this variance, we can use the classical Rao-Cramér inequality [2.50] if the measure μ_θ possesses the appropriate differentiability properties in θ. For the sake of simplicity we take the case of one parameter $\theta \in \mathbb{R}$. If we assume that there exists a second moment for the logarithmic derivative $\hat{\gamma}_\theta = \partial \ln \mu_\theta / \partial\theta$, which is the Radon-Nikodym derivative of measure $\mu_\theta' = \partial\mu_\theta/\partial\theta$, or

$$\mu_\theta\,(dx)\,\hat{\gamma}_\theta\,(x) = \partial\mu_\theta\,(dx)/\partial\theta, \tag{2.3.1.2}$$

we can easily obtain the inequality

$$\int (\hat{\theta}(x) - \theta)^2\,\mu_\theta(dx) \int (\hat{\gamma}_\theta(x) - \gamma_\theta)^2\,\mu_\theta(dx) \geqslant J_\theta^2, \tag{2.3.1.3}$$

where $J_\theta = \int \mu_\theta\,(dx)$ is the total intensity of the wave packet ψ_θ, and γ_θ is the mean value of $\hat{\gamma}_\theta$:

$$J_\theta\gamma_\theta = \int \hat{\gamma}_\theta(x)\,\mu_\theta(dx) = J_\theta' = 2\,\mathrm{Re}\,(\psi_\theta|\psi_\theta').$$

Inequality (2.3.1.3), which implies the inverse proportionality of the standard deviations $\sigma_{\hat{\theta}} \geqslant 1/\sigma_{\hat{\gamma}}$ or variances

$$\sigma_{\hat{\theta}}^2 = \int (\hat{\theta}(x) - \theta)^2\mu_\theta(dx)/J_\theta,\quad \sigma_{\hat{\gamma}}^2 = \int (\hat{\gamma}_\theta(x) - \theta)^2\ \mu_\theta(dx)/J_\theta \tag{2.3.1.4}$$

follows in an obvious manner from the Schwarz inequality if we allow for the fact that the right-hand side can be represented, in accordance with (2.3.1.1), in the form of the square of the scalar product

$$J_\theta = \int \hat{\theta}(x)\,\mu_\theta'(dx) - \theta J_\theta' = \int (\hat{\theta}(x) - \theta)\,(\hat{\gamma}_\theta(x) - \gamma_\theta)\,\mu_\theta(dx).$$

In a more general situation, where the estimated parameters $\theta = [\theta^i]_{i=1}^m$ are differentiable functions $\theta\,(\alpha)$ of unknown parameters $\alpha = [\alpha^k]_{k=1}^n$ of the density operators of mixed wave patterns $S\,(\alpha)$, we can easily obtain a matrix Rao-Cramér inequality that is invariant with respect to the choice of the state parameters:

$$R \geqslant D\sigma_{\hat{\gamma}\hat{\gamma}}^{-1}D^{\mathrm{T}}, \tag{2.3.1.5}$$

where $D = |\partial\theta^i/\partial\alpha^k|$ is the Jacobian of the $\alpha \mapsto \theta$ transformation, $R = \sigma_{\hat\theta\hat\theta}$ is the covariance matrix

$$R^{ik}(\alpha) = \int (\hat\theta^i(x) - \theta^i(\alpha))(\hat\theta^k(x) - \hat\theta^k(\alpha))\,\mu(\alpha, dx) \tag{2.3.1.6}$$

of unbiased estimators $\hat\theta^i(x)$ with respect to $\mu(\alpha, dx) = \operatorname{Tr} S(\alpha) M(dx)$,

$$\int \hat\theta^i(x)\,\mu(\alpha, dx) = \theta^i(\alpha) J(\alpha), \quad J(\alpha) = \operatorname{Tr} S(\alpha), \tag{2.3.1.7}$$

and $\sigma_{\hat\gamma\hat\gamma}$ is a similar covariance matrix for the logarithmic derivatives $\hat\gamma_k = \partial \ln \mu(\alpha)/\partial\alpha^k$, $k = 1, \ldots, n$. We will derive the inequality for the general noncommutative case.

2.3.1.2 Symmetric Bound

The lower bound of inequality (2.3.1.5) depends, naturally, on the choice of the representation determined by the method of measurement. By using the noncommutative analog of the logarithmic derivative introduced by Helstrom, we can obtain a more exact bound for the variances of the unbiased estimators that does not depend on the choice of representation.

If we assume that the family $\{S_\theta\}$ of the trace class density operators is strongly differentiable in θ in a certain region Θ, we can define a symmetric logarithmic derivative by the following equation:

$$\hat g_\theta S_\theta + S_\theta \hat g_\theta = 2S'_\theta. \tag{2.3.1.8}$$

It is easy to show (see [2.13]) that if $|\operatorname{Tr}(S'_\theta \hat x)|^2 \leqslant c \operatorname{Tr}(S_\theta \hat x^2)$ for every Hermitian operator $\hat x$, the solution to this equation exists and is unique, with $\operatorname{Tr}(S_\theta \hat g_0^2) < \infty$.

Let us consider the operator

$$\hat x = \int \hat\theta(x) M(dx), \quad \operatorname{Tr}(S_\theta \hat x) = \theta J_\theta,$$

determined by the unbiased estimator $\hat\theta$ for a fixed measurement M. Since

$$\int (\hat\theta(x) - \theta)^2 \mu_0(dx) = \operatorname{Tr}\left(S_0 \int (\hat\theta(x) - \theta)^2 M(dx)\right),$$

$$\operatorname{Tr}\left(S_0 \int (\hat\theta(x) - \hat x) M(dx) (\hat\theta(x) - \hat x) + (\hat x - \theta)^2\right) \geqslant \operatorname{Tr}[S_\theta(\hat x - \theta)^2],$$

it is sufficient to find the lower bound of the variance $\sigma^2_{\hat x}$ of operator $\hat x$.

By analogy with the commutative case we have

$$J_\theta = \mathrm{Tr}\,(\hat{x}S'_\theta - \theta J'_\theta) = \mathrm{Tr}\,[(\hat{x}-\theta)\,S'_\theta] = \frac{1}{2}\,\mathrm{Tr}\,[(\hat{x}-\theta)\,(\hat{g}_\theta S_\theta + S_\theta \hat{g}_\theta)]$$

$$= \frac{1}{2}\,\mathrm{Tr}\,[S_\theta\,((\hat{x}-\theta)\,(\hat{g}_\theta - \gamma_\theta) + (\hat{g}_\theta - \gamma_\theta)\,(\hat{x}-\theta))]$$

$$= \langle \hat{x}-\theta \,|\, \hat{g}_\theta - \gamma_\theta\rangle_+.$$

Thus, the total intensity J_θ is equal to the symmetrized scalar product with respect to S_θ of the operators $\hat{x} - \theta$ and $\hat{g}_\theta - \gamma_\theta$, where $\gamma_\theta = J'_\theta/J_\theta$ is the mean value of the operator $\hat{g}_\theta$ of the logarithmic derivative, and $J'_\theta = \mathrm{Tr}\,(S_\theta \hat{g}_\theta) = \mathrm{Tr}\,S'_\theta$. Applying the Schwarz inequality

$$|\,\langle \hat{x}-\theta \,|\, \hat{g}_\theta - \gamma_\theta\rangle_+\,|^2 \leqslant \langle \hat{x}-\theta \,|\, \hat{x}-\theta\rangle_+\,\langle \hat{g}_\theta - \gamma_\theta \,|\, \hat{g}_\theta - \gamma_\theta\rangle_+,$$

we arrive at the sought inequality:

$$\sigma_\theta^2 \geqslant \mathrm{Tr}\,[S_\theta\,(\hat{x}-\theta)^2]/J_\theta \equiv \sigma_{\hat{x}}^2 \geqslant J_\theta/\mathrm{Tr}\,[S_\theta\,(\hat{g}_\theta - \gamma_\theta)^2] \equiv 1/\sigma^2_{\hat{g}\,\theta}. \tag{2.3.1.9}$$

Thus, the variance of any unbiased estimator cannot be smaller than the inverse variance of the operator of the logarithmic derivative (2.3.1.8):

$$\sigma^2_{\hat{g}\,\theta} = \mathrm{Tr}\,[S_\theta\,(\hat{g}_\theta - \gamma_\theta)^2]/J_\theta. \tag{2.3.1.10}$$

A similar result can be obtained in the case where there are several parameters $\theta = [\theta^i]_{i=1}^m$ for the estimator $\hat{\theta}\,(x) = [\hat{\theta}^i\,(x)]_{i=1}^m$ satisfying the unbiasing conditions (2.3.1.1), which when met make matrix (2.3.1.5) the covariance matrix of estimators $\hat{\theta}^i$, and the mean square error at a fixed R_θ assumes the minimal value.

For the covariance matrix R_θ, Helstrom [2.14] has established the lower bound by assuming that the operator function $S_\theta = S\,(\theta)$ is differentiable and using the concept of the operators g_i of partial symmetrized logarithmic derivatives of the functions $S\,(\theta)$ in θ^i. He defined these operators by the following equations:

$$g_i S_\theta + S_\theta g_i = 2\,(\partial S_\theta/\partial\theta^i). \tag{2.3.1.11}$$

As in the classical case [2.50], this bound is defined by the matrix $G_\theta = \|\,G_{ik}\,(\theta)\,\|$ of the covariances of the solutions $g_i = g_i\,(\theta)$ of Eqs. (2.3.1.1). This matrix for noncommutative g_i is taken in symmetrized form

$$G_{ik}\,(\theta) = \mathrm{Tr}\,(S_\theta\,(g_i g_k + g_k g_i)/2) \tag{2.3.1.12}$$

(the mathematical expectations of Tr $(S_\theta g_i(\theta))$ are equal to zero). The corresponding inequality has the form

$$R_\theta \geqslant G_\theta^{-1}, \quad \theta \in \Theta \tag{2.3.1.13}$$

and is understood as the nonnegative definiteness of matrix $[R^{ik}(\theta) - G^{ik}(\theta)]$, where $G^{ik}(\theta)$ are the elements of the inverse matrix G_θ^{-1}: $G^{ij}(\theta)\, G_{jk}(\theta) = \delta^i_k$. Inequality (2.3.1.13) is the noncommutative analog of the Rao-Cramér inequality [2.50]. The matrix G_θ plays the role of a metric tensor locally defining the distance $\sigma(\theta, \theta + d\theta) = G_{ik}(\theta)\, d\theta^i\, d\theta^k$ in the parameters space Θ, similar to the Fisher information distance in classical statistics.

We now turn to a more general situation where the state parameters are not the measured parameters θ^i but other parameters $\alpha = \{\alpha^k, k = 1, \ldots, n\}$, $S = S(\alpha)$. The parameters θ^i are differentiable functions $\theta^i = \theta^i(\alpha)$ of the unknown parameters. The respective generalized Helstrom inequality (2.3.1.13) represents a bound for the covariance matrix $R = R(\alpha)$ of the estimators $\hat{\theta}^i$ in a form invariant with respect to the choice of the variables of state $S(\alpha)$,

$$R \geqslant DG^{-1}D^{\mathrm{T}}, \tag{2.3.1.14}$$

where $D = \| \partial\theta^i / \partial\alpha^k \|$, and $G = G(\alpha)$ is the covariance matrix (2.3.1.12) of the operators $g_k = g_k(\alpha)$ of symmetrized logarithmic derivatives of the operator function $S(\alpha)$ in α^k.

Inequality (2.3.1.14), which is equivalent to inequality (2.3.1.13) only if $m = n$ and matrix $D = D(\alpha)$ is nonsingular, can be verified by a line of reasoning similar to the one that will lead us to inequality (2.3.1.17) (see Section 2.3.2.5).

Inequality (2.3.1.14) can be reduced to the classical Rao-Cramér inequality only where the family $\{S(\alpha)\}$ is commutative. For noncommutative families other generalizations [2.22] of the Rao-Cramér inequality are possible, generalizations that are based on other definitions of logarithmic derivatives and that lead to other lower bounds for R differing from the invariant Helstrom bound $DG^{-1}D^{\mathrm{T}}$. For real-valued parameters α these generalizations may serve equally well as analogs of the Rao-Cramér inequality and coincide only if $\{S(\alpha)\}$ constitutes a commutative family, in which case they are reduced to the classical Rao-Cramér inequality. However, in the event of complex-valued parameters α a special invariant generalization of the Rao-Cramér inequality becomes especially important. It is based on the notions of right and left logarithmic derivatives and was suggested independently by Belavkin [2.31] and Yuen and Lax [2.49].

Let us assume that the parameters α^k are pairs (α_1^k, α_2^k) represented by complex numbers: $\alpha^k = \alpha_1^k + j\alpha_2^k$, $\alpha = \{\alpha^k\} \in \mathbb{C}^n$. The estimated parameters $\theta^i = \theta^i(\alpha, \bar{\alpha})$, $i = 1, \ldots, m$, are assumed

to be functions independently differentiable in α and $\bar{\alpha}$.[3] Let us define the non-Hermitian logarithmic derivatives of $S = S(\alpha, \hat{\alpha})$ thus:

$$Sh_k = \partial S/\partial\bar{\alpha}^k, \quad h_k^* S = \partial S/\partial\alpha^k, \quad k = 1, \ldots, n. \tag{2.3.1.15}$$

The operators $h_k = h_k(\alpha, \bar{\alpha})$ are called the right derivatives with respect to $\bar{\alpha}^k$ (and the operators h_k^* the left derivatives with respect to α) and have zero mathematical expectations. The covariance matrix

$$H_{ik}(\alpha, \bar{\alpha}) = \mathrm{Tr}\,[S(\alpha, \bar{\alpha})]\, h_i h_k^* \tag{2.3.1.16}$$

is Hermitian and, assuming it is nonsingular, defines a positive definite metric $ds^2 = H_{ik}\, d\bar{\alpha}^i\, d\alpha^k$ in a complex domain $\Theta \subset \mathbb{C}^n$ of the unknowns $\alpha \in \Theta$.

Suppose that a joint measurement of the parameters θ^i is described by a decomposition of unity that defines the estimator $\hat{\theta}$. This estimator is represented by a vector quantity that, in general, assumes complex values $x = \{x^i\} \in \mathbb{C}^m$, is represented by a conditional distribution $\mu(dx \mid \alpha, \bar{\alpha}) = \mathrm{Tr}\,[M(dx)\, S(\alpha, \bar{\alpha})]$, and satisfies the unbiasing condition $\langle\hat{\theta}^i\rangle = \theta^i(\alpha, \bar{\alpha})$. Then the mean square error of measurement is determined by the matrix $R = R(\alpha, \bar{\alpha})$ of covariances $R^{ik} = \langle(\hat{\theta}^i - \theta^i)(\hat{\theta}^k - \theta^k)^*\rangle$, for which the following inequality holds true:

$$R \geqslant DH^{-1}D^+, \tag{2.3.1.17}$$

where $D = D(\alpha, \bar{\alpha})$, as in (2.3.1.14), is the matrix of the derivatives $\partial\theta^i/\partial\alpha^k$, and D^+ is the respective Hermitian conjugate matrix. Even in the real case, that is $\hat{\theta}^i = \bar{\theta}^i$, inequality (2.3.1.17) leads to a lower bound that differs from the Helstrom bound (2.3.1.14). We will call the lower bound in (2.3.1.17) the right bound. Other bounds can also be considered, say, the left bound, which is based on the left logarithmic derivatives with respect to $\bar{\alpha}$. The proof of all such inequalities is similar to that of inequality (2.3.1.17), which is given in Section 2.3.2.5. The right bound in (2.3.1.17) is invariant under replacement of derivatives with respect to α^k by derivatives with respect to new variables $\beta^k = \beta^k(\alpha)$ only if the functions

[3] The derivatives $\partial/\partial\bar{\alpha}$ and $\partial/\partial\alpha$ are defined in terms of the partial derivatives $\partial/\partial\alpha_1$ and $\partial/\partial\alpha_2$ in the common manner:

$$\partial/\partial\alpha = \frac{1}{2}(\partial/\partial\alpha_1 + j\partial/\partial\alpha_2),$$

$$\partial/\partial\bar{\alpha} = \frac{1}{2}(\partial/\partial\alpha_1 - j\partial/\partial\alpha_2).$$

$\beta^k = \beta^k(\alpha)$ are analytic, that is, $\partial\beta^k/\partial\bar{\alpha}^i = 0$, and the matrix of derivatives $\partial\beta^k/\partial\alpha^i$ is nonsingular. Hence, the use of inequality (2.3.1.17) in invariant form $R \geqslant H^{-1}$, where, as in (2.3.1.13), we employ derivatives with respect to the estimated parameters θ^i (but, in contrast to (2.3.1.13), right derivatives rather than symmetrized are used), is inexpedient since the condition for the equivalence of these inequalities includes not only the condition that matrix D be nonsingular but the analyticity condition $\partial\theta^i/\partial\bar{\alpha}^k = 0$ as well (that is, the independence of functions $\theta^i(\alpha, \bar{\alpha})$ on $\bar{\alpha}$), which is not our initial assumption. A similar situation for complex-valued parameters α exists in the classical case.

2.3.2 Generalized Uncertainty Relations and Effective Measurements of Wave Patterns

In this section we will introduce the notion of canonical families of wave patterns for whose canonical parameters we will establish uncertainty relations that generalize the quantum mechanical uncertainty relations obtained in the one-dimensional case by Helstrom [2.13] and in the case of multidimensional Lie algebra by Belavkin [2.31]. We will then find the limit of accuracy in estimating the canonical Lie parameters of wave patterns and prove that such limits are exact only for canonical signals for which there are effective measurement or quasimeasurement procedures. The discourse will follow the scheme suggested in [2.31]; for examples of uncertainty relations for quantum systems the reader is advised to turn to [2.12, 2.13].

2.3.2.1 Canonical Families and Uncertainty Relations

In classical mathematical statistics an important role is played by canonical, or exponential, families of probability distributions, for which a special selection of parameters θ and α makes the Rao-Cramér bound exact. In Section 2.3.2.3 we will prove that in the noncommutative case a similar role is played by density operators of the form

$$S(\beta, \bar{\beta}) = \chi^{-1} e^{\beta^k \hat{x}_k^*} S_0 e^{\bar{\beta}^k \hat{x}_k}, \tag{2.3.2.1}$$

where the $\hat{x}_k$, $k = 1, \ldots, n$, are linearly independent operators in $\mathscr{H}$, which may be non-Hermitian ($\hat{x}_k^* \neq \hat{x}_k$) and may not commute with the conjugate operators ($\hat{x}_i\hat{x}_k^* \neq \hat{x}_k^*\hat{x}_i$), and $\chi = \chi(\beta, \bar{\beta})$ is the generating function of the moments of these operators in state S_0:

$$\chi(\beta, \bar{\beta}) = \operatorname{Tr} S_0 e^{\bar{\beta}^k \hat{x}_k} e^{\beta^k \hat{x}_k^*}, \tag{2.3.2.2}$$

which is finite ($\chi < \infty$) in a neighborhood of zero $\beta = 0$ of the complex space $\mathbb{C}^n$. The family of density operators (2.3.2.1) will be said to be canonical and the parameters β^k, canonically conjugate

to the $\hat{x}_k$. In contrast to the commutative case, even for Hermitian operators $\hat{x}_k$ it is meaningful to assume that the conjugate parameters β^k may have complex values.

Of special interest is the case, which has no classical analog, of canonical states (2.3.2.1) where the β^k are imaginary and the $\hat{x}_k$ are Hermitian. The parameters $\theta^k = \text{Im}\, \beta^k/(2\pi)$ acquire a dimensionality and meaning of quantities that are dynamically conjugate to the $\hat{x}_k$; for instance, if $\hat{x}$ is frequency, θ is time, if x is momentum, θ is displacement, if $\hat{x}$ is angular momentum, θ is the angle of rotation. The canonical states (2.3.2.1) at $\beta^k = 2\pi j\theta^k$ assume the form

$$S_\theta = e^{2\pi j\theta^k \hat{x}_k} S_0 e^{-2\pi j\theta^k \hat{x}_k} \tag{2.3.2.3}$$

and are unitary equivalent to state S_0, which corresponds to a zero value of θ. It has been established that if we put $\alpha = \beta$ and apply inequality (2.3.1.17) to the canonical family (2.3.2.3), we arrive at the exact formulation of the generalized uncertainty principle for any pair of dynamically conjugate quantities $\hat{\theta}^k$ and $\hat{x}^k$, where the first quantity in the pair may not correspond to the Hermitian operator that meaningfully describes in $\mathscr{H}$ the measurement of this quantity.[4]

Differentiating (2.3.2.1) with respect to $\bar{\beta}^k$ and comparing the result with (2.3.1.1), we get

$$h_i = e^{-\bar{\beta}^k \hat{x}_k} \chi \frac{\partial}{\partial \bar{\beta}^i} (\chi^{-1} e^{\bar{\beta}^k \hat{x}_k}) = \hat{x}_i(\bar{\beta}) - \theta_i, \tag{2.3.2.4}$$

where $\hat{x}_i(\bar{\beta}) = e^{-\bar{\beta}^k \hat{x}_k} \frac{\partial}{\partial \bar{\beta}^i} e^{\bar{\beta}^k \hat{x}_k}$, and $\theta_i = \frac{\partial}{\partial \bar{\beta}^i} \ln \chi = \text{Tr}\,[S\hat{x}_i(\bar{\beta})]$. Matrix (2.3.1.16), therefore, is the covariance matrix.

$$H_{ik} = \text{Tr}\,[S(\hat{x}_i(\bar{\beta}) - \theta_i)(\hat{x}_k(\bar{\beta}) - \theta_k)^*] = \frac{\partial^2 \ln \chi}{\partial \bar{\beta}^i \, \partial \beta^k} \tag{2.3.2.5}$$

of the operators $\hat{x}_i(\bar{\beta})$ analytic in $\bar{\beta}$ and coinciding with $\hat{x}_i$ at $\beta = 0$. The inequality (2.3.1.17) in the neighborhood of point $\beta = 0$,

---

[4] The Heisenberg uncertainty principle is usually proved only for such dynamically conjugate quantities described by noncommutative operators $\hat{p}$ and $\hat{q}$ that satisfy, say, the commutation relations $[\hat{p}, \hat{q}] = 1/2\pi j$. The proof employs the well-known scalar inequality $\langle(\hat{p} - \langle\hat{p}\rangle)^2\rangle \langle(\hat{q} - \langle\hat{q}\rangle)^2\rangle \geqslant |\langle[\hat{p}, \hat{q}]\rangle|^2/4$, which is valid for any pair of operators $\hat{p}$ and $\hat{q}$. Strengthening and matrix multidimensional generalization of this inequality in terms of the covariance estimations of an arbitrary family of noncommutative operators are suggested in [2.21].

therefore, can be written in the form of the uncertainty relation

$$R \geqslant DS^{-1}D^{+}, \tag{2.3.2.6}$$

which establishes the inverse proportionality between the matrix $S = \| S_{ik} \|$ of the covariances

$$S_{ik} = \mathrm{Tr}\,[S(\beta, \bar{\beta})(\hat{x}_i - \mu_i)(\hat{x}_k - \mu_k)^*] \tag{2.3.2.7}$$

of the operators $\hat{x}_i$, $\mathrm{Tr}\,[S(\beta, \bar{\beta})\,\hat{x}_i] = \mu_i$, and the covariance matrix R of the estimators $\hat{\theta}^i$ of the functions $\theta^i(\beta, \bar{\beta})$ of the conjugate parameters β and $\bar{\beta}$. At $\theta = \beta$, (2.3.2.6) assumes the canonical form $R \geqslant S^{-1}$.

In the scalar case ($n = 1$), $\hat{x}(\bar{\beta}) = \hat{x}$ for every β, and the uncertainty relation (2.3.2.6) transforms into strict inequality in the entire domain $\Theta \ni \beta$. Putting $\theta = \mathrm{Im}\,\beta/(2\pi)$ and allowing for the fact that $\partial\theta/\partial\beta = 1/4\pi j$, we obtain at $\hat{x}^k = \hat{x}$ the generalized uncertainty relation

$$(2\pi)^2 R_\theta \geqslant (1/4)\, S^{-1} \tag{2.3.2.8}$$

in terms of the variances $R_\theta = \langle(\hat{\theta} - \theta)^2\rangle_\theta$, with $S = \mathrm{Tr}\,[S_0(\hat{x} - \mu)^2]$, valid for any pair of dynamically conjugate quantities $\hat{\theta}$ and $\hat{x}$ defining the canonical family (2.3.2.3). For the quantum case of pure states $S_0 = |\psi_0)(\psi_0|$, the scalar inequality (2.3.2.8) has been derived from inequality (2.3.1.13) by Helstrom [2.11] by a complicated procedure for calculating the matrix elements of the operators of symmetrized logarithmic derivatives.

In the multidimensional case, when the operators $\hat{x}_k$ are pairwise commutative (but not necessarily with $\hat{x}_k^*$ and S_0), the situation is the same: $\hat{x}_k(\bar{\beta}) = \hat{x}_k$ for every $\beta \in \Theta$ and inequality (2.3.2.6) is strict. The averages μ_k and the covariances (2.3.2.7) at $\hat{x}_k = \hat{x}_k^*$ and $\beta^k = 2\pi j\theta^k$ are independent of θ and, therefore, coincide with the respective values at $\theta = 0$: $\mu_k = \mathrm{Tr}\,(S_0\hat{x}_k)$ and

$$S_{ik} = \mathrm{Tr}\,[S_0(\hat{x}_i - \mu_i)(\hat{x}_k - \mu_k)]. \tag{2.3.2.9}$$

The uncertainty relation (2.3.2.8) in this case acquires a matrix meaning: R_θ is the covariance matrix of estimators $\hat{\theta}^i$ of the canonical parameters belonging to a translation group in state S_θ, and S is the covariance matrix (2.3.2.9) of the generators $\hat{x}_k$ of this group, defining the lower bound $S^{-1}/16\pi^2$ for R_θ uniformly in every $\theta \in \mathbb{R}^n$.

We now take up the case of noncommutative $\{\hat{x}_k\}$. Suppose that the operators $\hat{x}_k$ form a Lie algebra:

$$\hat{x}_i\hat{x}_k - \hat{x}_k\hat{x}_i = C^j_{ik}\hat{x}_j, \tag{2.3.2.10}$$

where C^j_{ik} are structure constants. Here the operators $\hat{x}_i(\bar{\beta})$ in (2.3.2.4) are linear combinations of the generators $\hat{x}_i$:

$$\hat{x}_i(\bar{\beta}) = L^{-1}(-\bar{\beta})^j_i\,\hat{x}_j, \tag{2.3.2.11}$$

with $L(\xi) = \xi^k C_k (I - e^{-\xi C})^{-1}$ an n-by-n matrix that exists in a neighborhood $\mathcal{O} \in \mathbb{C}^n$ of zero $\xi = 0$, and the $C_k = \| C^j_{ik} \|$ are the generators of the adjoint representation of the commutation relations (2.3.2.10). Expressing the covariance matrix H of the operators (2.3.2.11) in terms of the covariances (2.3.2.7) of the generators $\hat{x}_i$, we get instead of (2.3.2.6) the inequality

$$R \geqslant DL^+S^{-1}LD^+, \tag{2.3.2.12}$$

where $L = L(-\beta)$. In the case of (2.3.2.3), the family of S_θ is unitary homogeneous with respect to the Lie group with Hermitian operators $\hat{x}_k$ and canonical parameters θ^k. Similarly to (2.3.2.8), we obtain a more general relationship

$$(2\pi)^2 R_\theta \geqslant \frac{1}{4} L^{\mathrm{T}}_\theta S^{-1} L_\theta, \tag{2.3.2.13}$$

where $L_\theta = \theta^i G_i (I - e^{-j\theta^k G_k})^{-1}$, $G_k = 2\pi j C_k$. Inequality (2.3.2.13) determines in the domain $\mathcal{O} \subset \mathbb{R}^n$ of convergence of the series expansion

$$(I - e^{-\theta^k G_k})^{-1} = \sum_{m=0}^{\infty} e^{-m\theta^k G_k}, \quad \theta = \{\theta^i\} \in \mathcal{O},$$

the lower bound of the mean square error in estimating the canonical parameters of the unitary representation $e^{2\pi j\theta^k \hat{x}_k}$ of a Lie group.

2.3.2.2 Effective Measurements and Quasimeasurements

In classical statistics, estimations whose covariance matrix assumes the minimal value and thus transforms, locally or globally, the Rao-Cramér inequality into an equality are known as effective (locally or globally, respectively). In the noncommutative case, the concept of effectiveness introduced by analogy with the classical concept loses its universality because the generalization of the Rao-Cramér inequality is not unique and the definitions of locally effective estimates [2.14, 2.22, 2.49] based on different variants of this generalization are not equivalent. For this reason

we distinguish between the effective measurements (or estimates) for which the invariant Helstrom bound (2.3.1.14) is attained and those for which the right bound (2.3.1.17) is attained, with the former called Helstrom effective and the latter, right effective. As we show below, the notion of right effectiveness is more universal: measurements that are Helstrom effective are right effective, but not vice versa. Let us first prove that Helstrom effective estimates exist globally for canonical families of density operators (2.3.2.1) if the operators $\hat{x}_k$ are Hermitian and pairwise commutative and if for the estimated parameters θ we take the derivatives $\theta_k = \partial \ln \chi/\partial\varkappa^k$ of the generating function $\chi(\varkappa) = \mathrm{Tr}\,(S_0 e^{\varkappa^k \hat{x}_k})$, where $\varkappa = \beta - \bar{\beta}$. The parameters θ_k selected in this manner coincide with the averages defined by the canonical subfamilies of the density operators,

$$S(\varkappa) = \chi^{-1}(\varkappa)\, e^{\varkappa^k \hat{x}_k/2} S_0 e^{\varkappa^k \hat{x}_k/2}, \tag{2.3.2.14}$$

with $\mathrm{Im}\,\beta^k = 0$:

$$\theta_k(\varkappa) = \mathrm{Tr}\,[S(\varkappa)\,\hat{x}_k] = \partial \ln \chi/\partial\varkappa^k. \tag{2.3.2.15}$$

Taking for parameters α^k the canonical parameters $\varkappa^k$ and differentiating the operator functions (2.3.2.14), we obtain the symmetrized logarithmic derivatives in $\varkappa^k$: $g_k = \hat{x}_k - \theta_k$. Thus, the covariances (2.3.1.12) coincide with the covariances of operators $\hat{x}_k$,

$$G_{ik} = \mathrm{Tr}\,[S(\varkappa)\,(\hat{x}_i - \theta_i)\,(\hat{x}_k - \theta_k)] = \frac{\partial^2 \ln \chi}{\partial\varkappa^i\,\partial\varkappa^k}, \tag{2.3.2.16}$$

which are equal to the derivatives $\partial\theta_i/\partial\varkappa^k$ defining matrix D in (2.3.1.14). Therefore, inequality (2.3.1.14) assumes the form $R \geqslant G$, or $\| R_{ik} - G_{ik} \| \geqslant 0$, where $R_{ik} = \langle(\hat{\theta}_i - \theta_i)(\hat{\theta}_k - \theta_k)\rangle$ are the covariance of the unbiased estimators $\hat{\theta}_k$: $\langle\hat{\theta}_k\rangle = \theta_k$. If for these estimators we take the results x_k of measurements of the observables $\hat{x}_k$ (which are compatible), then matrix R assumes the minimal value $R = G$. Thus, for the canonical families (2.3.2.14) with commutative operators $\hat{x}_k$ there exists a Helstrom effective measurement of the functions (2.3.2.15) of the canonical parameters $\varkappa_k$, which is the usual compatible measurement of the observables $\hat{x}_k$. The domain of this effectiveness, obviously, coincides with the domain $\Theta \subset \mathbb{R}^n$ for which $\chi(\varkappa) < \infty$, $\varkappa \in \mathbf{0}$. It has been established that the converse is true in the following sense.

Let the estimators $\hat{\theta}_k$ (i.e. the results of a measurement) have averages $\theta_k(\alpha)$ and covariances $R_{ik}(\alpha)$ that are differentiable in a certain domain, and let the matrices $R = [R_{ik}(\alpha)]$, $D = [\partial\theta_i/\partial\alpha^k]$

satisfy the conditions

$$\partial (R^{-1}D)^i_k/\partial \alpha^i = \partial (R^{-1}D^j_i)/\partial \alpha^k \tag{2.3.2.17}$$

(the regularity conditions). We can then introduce the canonical parameters $\varkappa^k = \varkappa^k(\alpha)$ defined uniquely by the derivatives $\partial \varkappa^i/\partial \alpha^k = 2\,(R^{-1}D)^i_k$ if we put $\varkappa^k(\alpha_0) = 0$ for a fixed α_0. It can easily be verified that for a family of density operators $S(\alpha)$ of canonical form (2.3.2.14), with $\varkappa^k = \varkappa^k(\alpha)$ differentiable functions possessing a nonzero Jacobian, the regularity conditions are met in an effective measurement at $\theta_k(\alpha) = \partial \ln (\chi(\varkappa(\alpha))/\partial \varkappa^k$ such that $R_{ik} = G_{ik}(\varkappa(\alpha))$ and $2\,(R^{-1}G)^i_k = \partial \varkappa^i/\partial \alpha^k$. Proof of the converse assertion that under the regularity conditions the global Helstrom effectiveness comes into play only for canonical families (2.3.2.14) is given in Section 2.3.2.5 for a more general situation involving complex variables.

Hence, we have proved the following

Theorem 2.3.2.1 *Under appropriate regularity conditions, inequality* (2.3.1.14) *is transformed into an equality in a certain domain* $\Theta \subset \mathbb{R}^n$ *if and only if the family of density operators* $S(\alpha)$ *has the canonical form* (2.3.2.14), *where* $\hat{x}_k$, $k = 1, \ldots, n$, *are commutative Hermitian operators in* $\mathscr{H}$, *and the canonical parameters* $\varkappa^k$, $k = 1, \ldots, n$, *are functions of parameters* α *defined by the equations*

$$\partial \ln \chi/\partial \varkappa^k = \theta_k(\alpha), \quad k = 1, \ldots, n.$$

2.3.2.3 The Theorem Regarding the Canonical Form of the Family of Density Operators

Suppose that in a certain domain $\Theta \subset \mathbb{C}^n$ the unbiased estimators $\hat{\theta}_k$ possess averages $\theta_k(\alpha, \bar{\alpha})$ and covariances $R_{ik}(\alpha, \bar{\alpha})$ that satisfy the regularity conditions (2.3.2.17), to which we adjoin the analyticity condition

$$\frac{\partial}{\partial \bar{\alpha}^k} R^{-1}D = 0. \tag{2.3.2.18}$$

Here we can introduce, as we did in Section 2.3.2.1, canonically conjugate parameters $\beta^k = \beta^k(\alpha)$ via the equations $\partial \beta^i/\partial \alpha^k = (R^{-1}D)^i_k$ and conditions $\beta^k(\alpha_0) = 0$ for a fixed $\alpha_0 \in \Theta$ with the functions $\beta^k(\alpha)$ being analytic in view of conditions (2.3.2.18).

Theorem 2.3.2.2 *Under the formulated regularity conditions, inequality* (2.3.1.17) *is transformed into an equality in a certain domain* $\Theta \subseteq \mathbb{C}^n$ *if and only if the family* $\{S(\alpha, \bar{\alpha}), \alpha \in \Theta\}$ *has the canonical form* (2.3.2.1), *with* $S_0 = S(\alpha_0, \bar{\alpha}_0)$ *for an* $\alpha_0 \in \Theta$, *the operators* $\hat{x}_k$, k $1, \ldots, n$, *simultaneously possess in* $\mathscr{H}$ *the property of the right*

proper decomposition of unity

$$\mathrm{I}=\int M\,(dx),\quad \hat{x}_k M\,(dx)=x_k M\,(dx),\quad x=\{x_k\}\in\mathbb{C}^n,\tag{2.3.2.19}$$

and the parameters β^k, $k=1,\ldots,n$, *are analytic functions* $\beta^k\,(\alpha)$ *determined by the equations*

$$\partial\ln\chi/\partial\bar{\beta}^k=\theta_k\,(\alpha,\bar{\alpha}),\quad \alpha\in\Theta.$$

Optimal estimation is then reduced to a quasimeasurement of the non-Hermitian operators $\hat{x}_k$ *described by the decomposition of unity* (2.3.2.19), *while the minimal mean square error is determined by the covariance matrix*

$$R_{ik}=\mathrm{Tr}\,[S\,(\hat{x}_i-\theta_i)\,(\hat{x}_k-\theta_k)^*].\tag{2.3.2.20}$$

Proof. Sufficiency is proved in the same way as in Section 2.3.2.1. Employing the fact of invariance of the right bound (2.3.1.17) under the analytic transformations $\alpha\to\beta$, we select for the variables α^k determining this bound the parameters β^k of the family of density operators (2.3.2.1). The elements $\partial\theta_i/\partial\beta^k$ of matrix D then coincide with the elements (2.3.2.5) of matrix H if we allow for the fact that $\theta_i=\partial\ln\chi/\partial\bar{\beta}^i$. Since according to (2.3.2.19) the operators $\hat{x}_k$ are commutative, $\hat{x}_i\hat{x}_k=\int x_i x_k M\,(dx)=\hat{x}_k\hat{x}_i$, we have $\theta_k=\mu_k$ and $H_{ik}=S_{ik}$, where the μ_k are the averages of the $\hat{x}_k$, and the S_{ik} are the covariances (2.3.2.7) of these operators. Hence, inequality (2.3.1.17) assumes the form $R\geqslant S$. What remains to be proved is that the measurement described by the decomposition of unity (2.3.2.19) leads to an estimation for which $R=S$ even when the operators are not commutative with the respective conjugates: $\hat{x}_i\hat{x}_k^*\neq\hat{x}_k^*\hat{x}_i$ (which occurs when decomposition (2.3.2.19) is non-orthogonal). To do this, it is sufficient to allow for the representation

$$\hat{x}_i=\int x M\,(dx),\quad \hat{x}_i\hat{x}_k^*=\int x_i\bar{x}_k M\,(dx),\quad x\in\mathbb{C}^n,\tag{2.3.2.21}$$

which is obtained by integrating the equations in (2.3.2.19) and the adjoint equation $M\,(dx)\,\hat{x}_k^*=\bar{x}_k M\,(dx)$. Thanks to (2.3.2.21) the covariances

$$R_{ik}=\int (x_i-\theta_i)\,(\bar{x}_k-\theta_k)\,\mathrm{Tr}\,[SM\,(dx)]\tag{2.3.2.22}$$

of the estimators $\hat{\theta}_k$ obtained as a result of a quasimeasurement of operators $\hat{x}_k$ coincide with the covariances S_{ik} of these operators,

which proves the effectiveness of this quasimeasurement for the density operators (2.3.2.1). The proof of the converse of Theorem 2.3.2.2 follows from the derivation of inequality (2.3.1.17) and will be discussed in Section 2.3.2.5.

2.3.2.4 Discussion and an Example

Thus, the condition of (right) effectiveness requires the existence of commutative operators possessing a joint right spectral decomposition and playing the role of sufficient statistics, which it is natural to call right effective. Here it is sufficient to restrict the discussion to the operators in the minimal subspace generated by the regions $S(\beta, \bar{\beta})\mathscr{H}$ with the density operators $S(\beta, \bar{\beta})$ for all $\beta(\alpha) \in \mathbb{C}^n$ for which $\alpha \in \mathcal{O}$. Even if we consider only the real values of parameters $\theta(\alpha, \bar{\alpha})$, optimal estimation can be described by non-Hermitian and noncommutative (with the conjugate) operators of the right-effective statistics and, therefore, may not be Helstrom effective. However, estimates that are Helstrom effective correspond, according to Theorem 2.3.2.1, to the particular case of right effectiveness where the $\hat{x}_k$ are Hermitian. If the operators $\hat{x}_k$ in (2.3.2.1) are non-Hermitian but commutative with the conjugate operators, the right-effective estimates also coincide with complexificated estimates, which are Helstrom effective. However, commutativity $\hat{x}_h\hat{x}_i^* = \hat{x}_i^*\hat{x}_h$ may not occur either.

Example. Let $\hat{x}_k = \varphi_k(\alpha)$, where the φ_k are entire functions $\mathbb{C}^r \to \mathbb{C}$, and let $A = \{A_i, i = 1, \ldots, r\}$ be the annihilation operators satisfying the commutation relations $A_iA_k - A_kA_i = 0$, $A_iA_k^* - A_k^*A_i = \delta_i^k$. It is well-known that the operators A have right eigenvectors $|\alpha) \in \mathscr{H}$, $\alpha \in \mathbb{C}^r$, that define the nonorthogonal decomposition of unity

$$I = \int |\alpha)(\alpha| \prod_{i=1}^{r} \pi^{-1}\operatorname{Re}\alpha_i d\operatorname{Im}\alpha_i, \qquad A_i|\alpha) = \alpha_i|\alpha).$$

It is obvious then that the operators $\hat{x} = \varphi(A)$ have a right proper decomposition of unity (2.3.2.19), where

$$M(dx) = dx \int \delta(x - \varphi(\alpha))\,|\alpha)(\alpha| \prod_{i=1}^{r} \pi^{-1} d\operatorname{Re}\alpha_i d\operatorname{Im}\alpha_i$$

(dx is the Lebesgue measure on $\mathbb{C}^n$, and $\delta(x - \varphi)$ is the Dirac delta function). Hence, optimal estimation of the parameters $\theta_h = \partial \ln \chi/\partial\bar{\beta}^k$ of the density operators (2.3.2.1) at $\hat{x} = \varphi(A)$ is right effective and can be reduced to a coherent measurement and estimation of $\theta = \varphi(\alpha)$ by the result α. In the particular case where $\varphi(\alpha)$

is a linear function and S_0 is a Gaussian state this fact has been established in [2.5].

Note that along with right and left lower bounds one can consider other combined bounds via the factorization $\theta = \theta_+ + \theta_-$ by appropriately defining the right derivatives with respect to θ_+ and the left derivatives with respect to θ_-. An interesting question arising in this connection is whether the class of effective estimations is exhausted by the estimations for which at least one such bound is attained.

Let us now consider the (right) effectiveness of estimating the parameters β^k of the canonical families (2.3.2.1). The inequality (2.3.1.17) corresponding to this case with $\theta^k = \beta^k$ has the form $R \geqslant H^{-1}$, where H is the matrix of derivatives (2.3.2.5). Without loss of generality, we can assume that $\mathrm{Tr}\,(\hat{x}_k S_0) = 0$.

Theorem 2.3.2.3 *The inequality $R \geqslant H^{-1}$ transforms into an equality if and only if the operators $\hat{x}_k$ in* (2.3.2.1) *possess a right joint decomposition of unity* (2.3.2.19), *the generating function of the moments* (2.3.2.2) *of these operators in state S_0 is Gaussian,* $\chi(\beta, \bar{\beta}) = \exp\{\bar{\beta}^i H_{ik} \beta^k\}$, *with H_{ik} independent of β and $\bar{\beta}$ and linear functions $y^k = (H^{-1})^{ki} x_i$ of the results x_k of joint quasimeasurement of observables $\hat{x}_k$ are selected for the estimators $\hat{\beta}^k$.*

Proof. Sufficiency of the above-formulated conditions for the existence of right-effective estimation is obvious: the fact that matrix H coincides with the covariance matrix S of operators $\hat{x}_k$ implies that the covariance matrix $R = H^{-1} S H^{-1}$ is equal to H^{-1}. Necessity follows from the necessary conditions of right effectiveness in Theorem 2.3.2.2, according to which the family $S(\beta, \bar{\beta})$ must have the form

$$S(\beta, \bar{\beta}) = \psi^{-1} e^{\theta_k \hat{y}^{k*}} S_0 e^{\bar{\theta}_k \hat{y}^k}, \qquad (2.3.2.23)$$

where $\psi = \mathrm{Tr}\,[S_0 e^{\bar{\theta}_k \hat{y}^k} e^{\theta_k \hat{y}^{k*}}]$, $\beta^k = \partial\theta_k \ln \psi / \partial\theta_k$, and the operators $\hat{y}^k$ possess the joint right decomposition of unity:

$$\mathrm{I} = \int M(dx), \quad \hat{y}^k M(dy) = y^k M(dy), \quad y = \{y^k\} \in \mathbb{C}^n.$$

Comparing (2.3.2.14) with (2.3.2.1), we conclude that $\theta_k \hat{y}^k = \bar{\beta}^k \hat{x}_k$, whence

$$\theta_k = H_{ki} \bar{\beta}^i, \quad \psi(\theta, \bar{\theta}) \quad \chi(\beta, \bar{\beta}) = \bar{\beta}^i H_{ik} \beta^k, \quad \hat{y}^k = (H^{-1})^{ki} \hat{x}_i.$$

The proof of Theorem 2.3.2.3 is complete.

2.3.2.5 Proof of Inequality (2.3.1.17)

We start with the one-dimensional case. Let $\hat{x}$ be a non-Hermitian operator in $\mathscr{H}$ for which

$$\operatorname{Tr}[\hat{x}S(\alpha, \bar{\alpha})] = \theta(\alpha, \bar{\alpha}). \tag{2.3.2.24}$$

Differentiating (2.3.2.24) with respect to α and employing definition (2.3.1.15) and the normalization condition $\operatorname{Tr} S(\alpha, \bar{\alpha}) = 1$, according to which $\operatorname{Tr}(Sh^*) = 0$, we obtain

$$\partial\theta/\partial\alpha = \operatorname{Tr}[S(\hat{x} - \theta) h^*].$$

Since the covariance $\operatorname{Tr}[S(\hat{x} - \theta) h^*]$ obeys the Schwarz inequality

$$|\operatorname{Tr}[S(\hat{x}-\theta) h^*]|^2 \leqslant \operatorname{Tr}[S(\hat{x}-\theta)(\hat{x}-\theta)^*] \operatorname{Tr}(Shh^*), \tag{2.3.2.25}$$

which reflects the fact that the determinant of the 2-by-2 covariance matrix $\operatorname{Tr}(Sh_ih_k^*)$, $i = 0, 1$, with $h_0 = (\hat{x} - \theta)$ and $h_1 = h$, is nonnegative, we can write

$$\operatorname{Tr}[S(\hat{x}-\theta)(\hat{x}^*-\bar{\theta})] \geqslant |\partial\theta/\partial\alpha|^2/\operatorname{Tr}(Shh^*). \tag{2.3.2.26}$$

This inequality, obviously, specifies the lower bound on the variance of the estimation of parameter $\theta = \theta(\alpha, \bar{\alpha})$ in the class of ordinary measurements described by normal operators $\hat{x}$. But since the normality condition, $\hat{x}\hat{x}^* = \hat{x}^*\hat{x}$, was not used in deriving (2.3.2.26), this bound is the lower one for the variance of any estimators $\hat{\theta}$ obtained as a result of arbitrary generalized measurements described in $\mathscr{H}$ by decompositions of unity $\mathrm{I} = \int M(dx)$, $x \in \mathbb{C}$ that may be nonorthogonal. Indeed, the nonnegative definiteness

$$(\hat{x}-x) M(dx)(\hat{x}-x)^* \geqslant 0 \quad (M \geqslant 0) \tag{2.3.2.27}$$

implies

$$\int |x-\theta|^2 M(dx) \geqslant (\hat{x}-\theta)(\hat{x}-\theta)^*, \tag{2.3.2.28}$$

where $\hat{x} = \int xM(dx)$, and $\theta = \operatorname{Tr}(S\hat{x})$. Taking the mathematical expectations of both sides of (2.3.2.27), allowing for the fact that the variance R of the estimator $\hat{\theta} = x$ is equal to $\operatorname{Tr} S \times \int |x - \theta|^2 M(dx)$, and combining the result with (2.3.2.26), we find that

$$R \geqslant \operatorname{Tr}[S(\hat{x}-\theta)(\hat{x}-\theta)^*] \geqslant |D|^2/H, \tag{2.3.2.29}$$

where $D = \partial\theta/\partial\alpha$, and $H = \mathrm{Tr}\,(Shh^*)$. This proves inequality (2.3.1.17) for the one-dimensional case.

The equality in (2.3.2.28) occurs if, first, the averages of both sides of (2.3.2.28) coincide and if, second, the Schwarz inequality transforms into an equality. Actually, the first requirement establishes an equality in (2.3.2.28). Specifically, we have the following

Lemma *Suppose that the ranges of values* $S(\alpha, \bar\alpha)\mathcal{H}$ *of the density operators from a certain family* $\{S(\alpha, \bar\alpha), \alpha \in \Theta\}$ *generate the entire space* $\mathcal{H}$. *Then the fact that* $\mathrm{Tr}\,(SA) = 0$ *for every nonnegative definite operator* A *in* $\mathcal{H}$ *and all* $\alpha \in \Theta$ *implies that* $A = 0$.

Proof. It is sufficient to prove that in $\mathcal{H}$ there is no vector χ of the form $\chi = S^{1/2}\psi$ for which $(\chi \mid A \mid \chi) \neq 0$. But this follows from the well-known inequality

$$\mathrm{Tr}\,(S^{1/2}AS^{1/2}) \geqslant (\psi \mid S^{1/2}AS^{1/2} \mid \psi),$$

which is true for every nonnegative A at $(\psi \mid \psi) = 1$.

Applying this result to the operator A that is equal to the difference between the right- and left-hand sides of (2.3.2.28), we find that under the lemma's hypothesis the equality in (2.3.2.28) occurs only if

$$(\hat{x} - x)\,M(dx)\,(\hat{x} - x)^* = 0, \quad \text{or} \quad \hat{x}M(dx) = xM(dx).$$

This proves that right-effective estimation in a certain region $\Theta \ni \alpha$ exists if there is an operator of minimal sufficient statistics, $\hat{x}$, possessing a right proper decomposition of unity in the subspace generated by the subspaces $S(\alpha, \bar\alpha)\mathcal{H}$. In the real case, $x \in \mathbb{R}$, such an operator is obviously Hermitian.

The second requirement for equality to occur in (2.3.2.28) is equivalent to the condition of linear dependence, $Sh = \bar\lambda S(x - 0)$, where $\lambda = D/R$, if the first condition for equality in (2.3.2.28) is met. Extending this condition over the entire region $\Theta \ni \alpha$ in which the analyticity condition (2.3.2.8), $\partial\lambda/\partial\bar\alpha = 0$, is assumed to hold true, we arrive at the equation $\partial S/\partial\bar\alpha = \bar\lambda S(\hat{x} - \theta)$ in $S = S(\alpha, \bar\alpha)$. Its solution combined with the boundary condition $S(\alpha_0, \bar\alpha_0) = S_0$ has the canonical form (2.3.2.1), where $\beta(\alpha) = \int_{\alpha_0}^{\alpha} \lambda(\alpha)\, d\alpha$ is an analytic function, and $\hat{x}$ is the operator of right-effective statistics. This proves that in the one-dimensional case the existence of right-effective estimation requires that the density operators $S(\alpha, \bar\alpha)$ be canonical. This condition is formulated in Theorem 2.3.2.2. For the real case, $\hat{x}^* = \hat{x}$, this fact is proved in Theorem 2.3.2.1.

The multidimensional generalization can be carried out if for $\hat{x} - \theta$ and h we take the sums $(\hat{x}^i - \theta^i)\,\bar\eta_i$ and $h_k\bar\xi^k$, where η_i, $i =$

$1, \ldots, m$ and ξ^k, $k = 1, \ldots, n$, are complex numbers. If we allow for the fact that here $\mathrm{Tr}\,[S\,(\hat{x} - \theta)\,h^*] = \bar{\eta}_i\,(\partial\theta^i/\partial\alpha^k)\,\xi^k$, then from (2.3.2.25) at $\xi^k = (H^{-1}D^+)^{ki}\,\eta_i$ we arrive at the inequality

$$R^{ik}\bar{\eta}_i\eta_k \geqslant \mathrm{Tr}\,[S\,(\hat{x}^i - \theta^i)\,(\hat{x}^k - \theta^k)^*\,\eta_i\eta_k] \geqslant (DH^{-1}D^+)^{ik}\,\bar{\eta}_i\eta_k$$

valid for an arbitrary $\hat{x}^i$ for which $\mathrm{Tr}\,(S\hat{x}^i) = \theta^i$. Putting $\hat{x}^i = \int x^i M\,(dx)$, where $\int M\,(dx) = I$, $x \in \mathbb{C}^m$, is the decomposition of unity describing the estimator $\hat{\theta}^i = x^i$, and applying inequality (2.3.2.28) with $\hat{x} = \hat{x}^i\bar{\eta}_i$ and $\theta = \theta^i\bar{\eta}_i$, we obtain for the matrix R of covariances of $\hat{\theta}^i$ the first inequality in (2.3.2.29), which in view of the arbitrariness of η_i yields (2.3.1.17).

Inequality (2.3.2.29) transforms into an equality at $\alpha \in \mathcal{O}$ only when $\hat{x}^i M(dx) = x^i M\,(dx)$ and $\partial S/\partial\bar{\alpha}^k = \bar{\lambda}_{ki} S\,(x^i - \theta^i)$, where $\lambda_{ik} = (R^{-1}D)_{ik}$, whence, if we allow for the regularity conditions λ_{ik}, we arrive at (2.3.2.1).

2.3.3 Optimal and Covariant Estimation of the Parameters of Wave Patterns

In this section we will consider the necessary and sufficient conditions for the optimality of measuring sound and visual patterns by the criterion of mean square error in parameter estimation and by the maximal intensity criterion. To avoid substantiation of the operator integrals involved in the discussion (this is done in [2.32]), we interpret them as operator-valued Radon measures. The solution to the optimal measurement problem will be found for homogeneous families of wave patterns for which it coincides with optimal covariant measurements of the corresponding parameters of quantized fields, with the latter measurements introduced in [2.29].

2.3.3.1 Optimal Measurements

The problems of optimal estimation of continuous wave parameters constitute essentially multialternative problems with an infinite-dimensional solution space (or manifold) X. Without loss of generality, we can assume that the information parameter space Θ coincides with X equipped with measure $d\lambda$. Let us assume that a wave signal, which in general is described by a density operator S, depends in a continuous manner on real- or complex-valued random parameters $\theta = (\theta_1, \ldots, \theta_n)$, $S = S_\theta$, having a given a priori distribution $P\,(d\theta)$. The deviation of the estimate $x \in X$ from θ is penalized by an integrable cost function $c_x\,(\theta)$ of the form, say, $(x - \theta)^2$. On X we must find an optimal quasimeasurement that (a) is described by an operator-valued measure $M\,(dx)$, (b) determines the decomposition of unity in the Hilbert space $\mathscr{H}$, and (c) minimizes

the mean estimation cost

$$\langle c\rangle = \int\int \mu_\theta(dx)\, c_x(\theta)\, P(d\theta) = \int \operatorname{Tr} R_x M(dx),$$

where $\mu_\theta(dx) = \operatorname{Tr} M(dx) S_\theta$ is the observed intensity distribution on X for a given θ, and $R_x = \int c_x(\theta)\, S_\theta P(d\theta)$ is the operator of the mean cost $x \in X$. We will now formulate the necessary and sufficient conditions for the optimality of solution M^0 to this extremal problem, which in [2.32] were introduced to estimate the parameters of quantum states. This will be done in a manner similar to that of Theorem 2.2.3.1:

Theorem 2.3.3.1 *The lower bound*

$$\inf_{M\geqslant 0}\left\{\int \langle R_x, M(dx)\rangle \,\middle|\, \int M(dx) = \mathrm{I}\right\} \tag{2.3.3.1}$$

is attained on measure M^0 if and only if for almost all $x \in X$ there exists a minorant operator $\Lambda^0 \leqslant R_x$ such that

$$(R_x - \Lambda^0)\, M^0(dx) = 0 \quad \forall x \in X. \tag{2.3.3.2}$$

The operator Λ^0 is a trace class operator, or $\operatorname{Tr}\Lambda^0 = \langle\Lambda^0, \mathrm{I}\rangle < \infty$, *that determines the solution to the duality problem*

$$\sup_\Lambda \{\langle\Lambda, \mathrm{I}\rangle \mid \Lambda \leqslant R_x,\ x \in X\} \tag{2.3.3.3}$$

for which conditions (2.3.3.2) *are also necessary and sufficient* (*if we allow for the fact that $M^0 \geqslant 0$ and $\int M^0(dx) = E$*).

Proof. For the proof of this theorem as well as for the existence conditions for a solution see [2.32].

Allowing for the fact that the operators $M(dx)$ can be decomposed into operators of the form $|\chi_x)(\chi_x|\, d\lambda(x)$, where the χ_x are the generalized elements of space $\mathcal{H}$, we find that the problem of optimal estimation of wave parameters will be solved if and only if we can find a family of reference waves, $\{\chi_x\}$ satisfying the completeness condition

$$\int |\chi_x)(\chi_x|\, d\lambda(x) = \mathrm{I} \tag{2.3.3.4}$$

and a Hermitian operator Λ for which

$$R_x - \Lambda \geqslant 0, \quad (R_x - \Lambda)\,\chi_x = 0, \quad x \in X. \tag{2.3.3.5}$$

Note that, in contrast to problems of signal discrimination, in problems of parameter estimation the commutative case $R_x R_{x'} = R_{x'} R_x$, which can be reduced to the classical case, is of no practical interest and will not be discussed here.

The solution of problem (2.3.3.5) poses no fundamental difficulties in the case of a single unknown real-valued parameter θ ($\chi = \mathbb{R}^1$) and a quadratic penalty function

$$C_x(\theta) = (x - \theta)^2.$$

The mean estimation cost operator

$$R_x = \int (x-\theta)^2 S_\theta P(d\theta)$$

in the case of (2.3.3.5) can be represented, via three Hermitian operators

$$R^{(k)} = \int \theta^k S_\theta P(d\theta), \quad k = 0, 1, 2, \tag{2.3.3.6}$$

in the form

$$\begin{aligned} R_x &= x^2 R^{(0)} - 2x R^{(1)} + R^{(2)} \\ &= (\hat{x} - x) R^0 (\hat{x} - x) + R^{(2)} - \hat{x} R^{(0)} x, \end{aligned}$$

where $\hat{x}$ is an operator satisfying the equation

$$\hat{x} R^{(0)} + R^{(0)} \hat{x} = 2R^{(1)}. \tag{2.3.3.7}$$

If we now put

$$\Lambda = R^{(2)} - \hat{x} R^{(0)} \hat{x}$$

and for χ_x take the complete orthogonal system of generalized eigenvectors determining the spectral decomposition of the Hermitian operator $\hat{x}$,

$$(\hat{x} - x)\chi_x = 0, \quad \hat{x} = \int x \mid \chi_x)(\chi_x \mid dx,$$

the conditions (2.3.3.5) are satisfied in an obvious manner:

$$(\hat{x} - x) R^{(0)} (\hat{x} - x) \geqslant 0, \quad (\hat{x} - x) R^{(0)} (\hat{x} - x)\ \chi_x = 0.$$

Thus, the solution of the parameter estimation problem by criterion (2.3.3.5) is reduced to measuring operator $\hat{x}$ satisfying Eq. (2.3.3.7). The result of such a measurement, x, leads to the minimal error $\langle c \rangle = \mathrm{Tr}\, \Lambda^0$ equal to the a posteriori variance

$$\sigma^2 = \mathrm{Tr}\,(R^{(2)} - \hat{x} R^{(0)} x).$$

As an example, let us consider the estimation of the amplitude of a coherent signal of known shape received against a background of Gaussian noise. The density operator of the corresponding mode

has the Gaussian form

$$S(\theta) = \int |\alpha)(\alpha| \bar{n}^{-1} \exp\left\{-\frac{|\alpha-\theta|^2}{\bar{n}}\right\} \pi^{-1} d\,\mathrm{Re}\,\alpha d\,\mathrm{Im}\,\alpha, \tag{2.3.3.8}$$

where θ is the amplitude, which assumes real values, $\theta \in \mathbb{R}^1$. We assume that amplitude θ has a Gaussian a priori density

$$p(\theta) = (2\pi\bar{s})^{-1/2} \exp\{-\theta^2/2\bar{s}\},$$

where $\bar{s}$ is the a priori variance, $\langle\theta^2\rangle = \bar{s}$. It is then easy to find, via the formulas of Gaussian integration, the operators (2.3.3.6), which define the mean decision cost operator

$$R_x = \left(x - \frac{2\bar{s}}{2\bar{s}+\bar{n}+1/2} Q\right) S \left(x - \frac{2\bar{s}}{2\bar{s}+\bar{n}+1/2} Q\right) + \frac{2\bar{s}(\bar{n}+1/2)}{2\bar{s}+\bar{n}+1/2} S.$$

Here $Q = \int \mathrm{Re}\,\alpha\,|\alpha)(\alpha|\,\pi^{-1}\,d\,\mathrm{Re}\,\alpha\,d\,\mathrm{Im}\,\alpha$ is the operator of the "coordinate" of the harmonic oscillator representing this mode, and

$$S = \int |\alpha)(\alpha| \left((2\bar{s}+\bar{n})\bar{n}\right)^{-1/2} \exp\left\{-\frac{(\mathrm{Re}\,\alpha)^2}{2\bar{s}+\bar{n}} - \frac{(\mathrm{Im}\,\alpha)^2}{\bar{n}}\right\} \pi^{-1} d\,\mathrm{Re}\,\alpha d\,\mathrm{Im}\,\alpha$$

is the density operator, with $\int S_\theta p(\theta)\,d\theta = R^{(0)}$. Hence, optimal estimation of the amplitude of a Gaussian signal is reduced to measuring the coordinate operator Q, whose result q determines the optimal estimate

$$x = 2\bar{s}q/(2\bar{s} + \bar{n} + 1/2)$$

with a minimal mean square error

$$\sigma^2 = 2\bar{s}\,(\bar{n} + 1/2)\,(2\bar{s} + \bar{n} + 1/2).$$

2.3.3.2 The Optimal Estimation Problem in the Multidimensional Case

In this case ($n > 1$) even for the quadratic quality criterion

$$C_x(0) = \sum_{j=1}^{n} (x_j - \theta_j)^2$$

the general solution to problem (2.3.3.5) is unknown. Only in the particular case where the operators $\hat{x} = \{\hat{x}_j\}$ obeying Eq. (2.3.3.7), with

$$R^k = R_j^k \equiv \int \theta_j^k S_\theta p(\theta) d\theta_1 \ldots d\theta_n, \quad k = 0, 1, 2,$$

commute with each other ($x_j x_i = x_i x_j$), optimal estimation is reduced, obviously, to joint measurement of these operators.

In general, a good estimate of parameters θ_j can be obtained by an indirect measurement of noncommutative operators $\{\hat{x}_i\}$ (see Section 2.3.2.2). However, this estimate is not necessarily optimal, even if the indirect measurement is ideal.

For an example let us take the complex-valued one-dimensional case ($X = \mathbb{C}^1$), which can also be interpreted as the real-valued two-dimensional:

$$C_x(\theta) = |x - \theta|^2 = (\mathrm{Re}\,(x - \theta))^2 + (\mathrm{Im}\,(x - \theta))^2.$$

An exact solution to the problem of optimal estimation of a single parameter θ has been obtained in [2.11] for this case of a quadratic penalty function, a Gaussian state S_θ of the form (2.3.3.8), with $\theta \in \mathbb{C}^1$, and a Gaussian a priori probability density

$$p(\theta) = \bar{s}^{-1} \exp\{-|\theta|^2/\bar{s}\}, \quad \bar{s} = \langle |\theta|^2 \rangle. \tag{2.3.3.9}$$

The density $p(\theta)$ is normalized with respect to

$$d\lambda(\theta) = \pi^{-1} d\,\mathrm{Re}\,\theta\, d\,\mathrm{Im}\,\theta.$$

In this case, by the standard formulas of Gaussian integration, we can easily find the mean decision cost operator

$$R_x = \left(x^* - \frac{\bar{s}}{\bar{n} + \bar{s} + 1} A^*\right) S \left(x - \frac{\bar{s}}{\bar{n} + \bar{s} + 1} A\right) + \frac{\bar{s}(\bar{n} + 1)}{\bar{s} + \bar{n} + 1} S,$$

where $A = \int \alpha \mid \alpha \rangle \langle \alpha \mid d\mu(\alpha)$ is the quantum annihilation operator, and

$$S = \int |\alpha)(\alpha| \frac{1}{\bar{s} + \bar{n}} \exp\left\{-\frac{|\alpha|^2}{\bar{s} + \bar{n}}\right\} d\lambda(\alpha)$$

is the density operator $S = \int S_\theta p(\theta)\, d\lambda(\theta)$. Assuming that

$$\Lambda = \frac{\bar{s}(\bar{n} + 1)}{\bar{s} + \bar{n} + 1} S, \quad \chi_x = c^{-1} \left| \left(1 + \frac{\bar{n} + 1}{\bar{s}}\right) x\right),$$

where $|\alpha)$, $\alpha = (1 + (\bar{n} + 1)/\bar{s})\, x$, are coherent vectors, and $c = \bar{s}/(\bar{s} + \bar{n} + 1)$ is a coefficient that can be found from condition (2.3.3.4) if we allow for the completeness of coherent states

$$\int |\alpha)(\alpha|\, d\lambda(\alpha) = I,$$

and allowing for the equation $A\,|\alpha) = \alpha\,|\alpha)$, we find that conditions (2.3.3.5) are met:

$$(x^* - cA^*) S (x - cA) \geqslant 0, \quad (x^* - cA^*) S (x - cA) |c^{-1}x) = 0. \tag{2.3.3.10}$$

Thus, optimal estimation in the one-dimensional complex-valued quadratic-Gaussian case is reduced to a coherent measurement describing an ideal proper indirect measurement of the annihilation operator A whose result α determines the estimate

$$x = \frac{\bar{s}}{\bar{s}+\bar{n}+1}\,\alpha$$

with a minimal error

$$\sigma^2 = \frac{\bar{s}\,(\bar{n}+1)}{\bar{s}+\bar{n}+1}.$$

This error is equal to the error of the appropriate classical problem of estimation in a Gaussian linear channel with a noise intensity of $\bar{n}+1$. The quantity $\bar{n}$ (the mean number of the noise quanta) is determined by the noise proper in the wave channel, while the unity corresponds to the "effective noise" thanks to the inaccuracy in the ideal indirect measurement. The measurement noise of unit intensity can be interpreted as the noise produced by an ideal wave amplifier or as the noise produced by an ideal optical heterodyne.

2.3.3.3 Optimal Measurement of Wave States

Let X be a set of hypotheses concerning the states of a wave field, and $\{\hat{R}_x,\ x \in X\}$ the respective decomposable family of density operators $R_x = \bigoplus_n R_x^{(n)}$ in the Hilbert space $\mathcal{H} = \bigoplus_n \mathcal{H}^{(n)}$. Then the set of measurements described by operator-valued measures M on X ($M(\cdot) \geqslant 0$, $\int M(dx) = I$) of the decomposable form $M(dx) = \bigoplus_n M^{(n)}\,dx$ is sufficient. The optimal strategy is described by the family of operator-valued measures $M^{(n)}$ on X ($M^{(n)}(\cdot) \geqslant 0$, $\int M^{(n)}(dx) = I^{(n)}$) defined independently in $\mathcal{H}^{(n)}$ for every n by the conditions

$$(R_x^{(n)} - \Lambda^{(n)})\,M^{(n)}(dx) = 0, \quad R_x^{(n)} \leqslant \Lambda^{(n)} \quad \forall\, x \in X \tag{2.3.3.11}$$

(the maximum intensity criterion). Here $\Lambda^{(n)}$ are Hermitian trace class operators in $\mathcal{H}^{(n)}$ that are nonnegative (for $R_x^{(n)} \geqslant 0$) and can be represented in the form

$$\Lambda^2 = \int R_x^{(n)} M^{(n)}(dx)\, R_x^{(n)}.$$

Problem (2.3.3.11) is incomparably simpler than the general problem of optimal discrimination of a (indecomposable) family $\{R_x\}$ and for every n has a finite-dimension of space $\mathcal{H}^{(n)}$ if the signal space $\mathcal{X}$ is finite-dimensional. In what follows, the index n will be dropped.

Let $\mathcal{U}_x = R_x\mathcal{H}$ be the range of values of operators R_x in $\mathcal{H}$, let $\mathcal{U}(dx)$ be their algebraic sum for all $x \in dx$, and let $\mathcal{U} = \int \mathcal{U}(dx)$ be the sum of all the subspaces $\mathcal{U}_x \subset \mathcal{H}$. Each nonnegative operator R_x can be represented in the form $R_x = \psi_x\psi_x^+$, where ψ_x is the operator from $\mathcal{U}_x$ into $\mathcal{U}$. The following conjectures are multidimensional generalizations of the appropriate assertions of Theorem 2.2.2.2 (for a discrete set X).

Theorem 2.3.3.2 (1) *Subspace $\mathcal{U}$ is sufficient for solving problem* (2.3.3.11). *Every operator Λ satisfying conditions $\Lambda \geqslant R_x$ $\forall\, x \in X$ for $R_x \geqslant 0$ has an inverse Λ^{-1} in $\mathcal{U}$.*

(2) *The solution to problem* (2.3.3.11) *in the sufficient space $\mathcal{U}$ has the form*

$$M(dx) = \Lambda^{-1}\psi_x\hat{\mu}(dx)\psi_x^+\Lambda^{-1}, \text{ where } \Lambda = \left(\int \psi_x\hat{\mu}(dx)\psi_x^+\right)^{1/2} \tag{2.3.3.12}$$

and $\hat{\mu}$ is a measure on X whose values $\hat{\mu}(dx)$ are nonnegative operators in $\mathcal{U}_x$ defined by the conditions

$$(\psi_x^+\Lambda^{-1}\psi_x - I_x)\,\hat{\mu}(dx) = 0, \quad \psi_x^+\Lambda^{-1}\psi_x \leqslant I_x \quad \forall x \in X \tag{2.3.3.13}$$

(I_x is the identity element in $\mathcal{U}_x$).

(3) *If the subspace $\mathcal{U}(dx)$ does not intersect with the sum $\mathcal{U}(\overline{dx})$ of all the remaining subspaces $\mathcal{U}_y$, $y \notin dx$, the operator $\hat{\mu}(dx)$ is strictly positive in $\mathcal{U}_x$ and is defined by the condition $\psi_x^+\Lambda^{-1}\psi_x = I_x$, $x \in X$.*

2.3.3.4 Fields with Group Symmetry

Equations (2.3.3.13) are considerably simpler than Eqs. (2.3.3.11) since the dimensionality of each operator equation in (2.3.3.13) is equal to rank $r(R_x)$. For the case where $r(R_x) = 1$ the solution has been found [2.4] under the condition that the square root of the correlation matrix $[\psi_x^+\psi_y]$ has equal diagonal elements. An analog of this condition in the general case where $r(R_x) \geqslant 1$ is the condition of group (say, cyclic in [2.5]) symmetry of the family $\{R_x\}$.

Let X be a homogeneous set with respect to a group G, that is, group G acts on X transitively, and let $U(g)$, $g \in G$, be a unitary representation of G in $\mathcal{H}$. The family $\{R_x, x \in X\}$ is said to be G-homogeneous (or G-invariant) if X is a homogeneous set with respect to group G and $U(g)\, R_{gx}^{-1} U^+(g) = R_x$.

Let G be a finite compact or locally compact group, dg be the left Haar measure on G, the family $\{R_x\}$ be homogeneous and continuous in $U(g)$. The following conjectures are true:

Theorem 2.3.3.3 (1) *The sufficient space $\mathcal{U}$ is a subspace in $\mathcal{H}$ cyclically generated by the family $\{U(g), g \in G\}$ over $\mathcal{U}_0 = R_{x_0}\mathcal{H}$,*

where x_0 is any element belonging to X. The operators R_x in $\mathfrak{U}$ can be represented in the form

$$R_x = U(g)\,\psi\psi^+U^+(g) \quad \forall\, g \in G,$$

where $U(g)$ is a subrepresentation induced in $\mathfrak{U} \subset \mathcal{H}$, ψ an operator from $\mathfrak{U}_0$ into $\mathfrak{U}$, and G_x the left coset $G_x = \{g: gx_0 = x\}$ over the stationary subgroup $G_0 = G_{x_0}$ of element x_0.

(2) *The optimal strategy* (2.3.3.12) *has the covariant form*

$$M(dx) = \int_{G(dx)} U(g)\,\Lambda^{-1}\psi\hat{\mu}\psi^+\Lambda^{-1}U^+(g)\,dg, \tag{2.3.3.14}$$

where $\Lambda = \left(\int U(g)\,\psi\hat{\mu}\psi^+U^+(g)\,dg\right)^{1/2}$ is the G-invariant, $G(dx) = \bigcup_{x\in dx} G_x$ the union of the G_x over all $x \in dx$, and $\hat{\mu}$ a non-negative operator in $\mathfrak{U}_0$ satisfying the conditions

$$(\psi^+\Lambda^{-1}\psi - \hat{I})\,\hat{\mu} = 0, \quad \psi^+\Lambda^{-1}\psi \leqslant \hat{I}. \tag{2.3.3.15}$$

(3) *If the representation $U(g)$ in $\mathfrak{U}$ is topologically irreducible, then operator Λ is a multiple of the identity element $\hat{I}$ of space $\mathfrak{U}_0$: $\Lambda = \lambda I$ and operator $\hat{\mu}$ is proportional to the proper projector $\hat{\pi}$ of operator $\hat{R} = \psi^+\psi$ corresponding to its maximal eigenvalue λ: $\hat{\mu} = \mu\hat{\pi}$, $(\hat{R} - \Lambda)\,\hat{\pi} = 0$.*

The proportionality factor can be made equal to unity by appropriately renormalizing dg. For the particular case where G is finite and its action on X is effective this result was obtained earlier in [2.24]. If G is an Abelian group, the case is trivial and can be of no interest.

(4) *Let $U_\omega(g)$, $\omega \in \Omega$, be the field of nonequivalent irreducible representations $U_\omega(g)$ in space $\mathcal{H}_\omega$, and $d\omega$ the Plancherel measure. Then* (2.3.3.15) *can be represented in the form*

$$\int_\Omega \mathrm{Tr}_{\mathcal{H}_\omega}(\hat{R}_\omega\hat{\mu})^{1/2}\,d\omega = \hat{\mu}, \quad \int_\Omega \mathrm{Tr}_{\mathcal{H}_\omega}(\hat{R}_\omega\hat{\mu})^{-1/2}\,\hat{R}_\omega\,d\omega \leqslant \hat{I}, \tag{2.3.3.16}$$

where $\hat{R}_\omega = \int \hat{r}(g)\,U_\omega(g)\,dg$ is the Fourier transform of the operator correlation function $\hat{r}(g) = \psi^+U^+(g)\,\psi$. If the family $\{U_\omega(\cdot)\}$ is discrete and $d\omega$ is the dimensionality of representations $U_\omega(g)$ (formally, if $\mathcal{H}_\omega$ is infinite-dimensional), then conditions (2.3.3.16) *assume the form*

$$\mathrm{Tr}_{\mathcal{H}_\omega}(\hat{R}_\omega\hat{\mu})^{1/2}\,d\omega = \hat{\mu}, \quad \mathrm{Tr}_{\mathcal{H}_\omega}(\hat{R}_\omega\hat{\mu})^{-1/2}\,\hat{R}_\omega\,d\omega \leqslant \hat{I}. \tag{2.3.3.17}$$

If $\mathcal{U}_0$ is one-dimensional, (2.3.3.16) and (2.3.3.17) lead us to the following solutions:

$$\mu = \Big(\int_{\Omega} \mathrm{Tr}_{\mathcal{H}_\omega} \hat{R}_\omega^{1/2} \, d\omega \Big)^2, \quad \mu = \Big(\sum_{\Omega} \mathrm{Tr}_{\mathcal{H}_\omega} \hat{R}_\omega^{1/2} \, d\omega \Big)^{1/2},$$

which were found in [2.25] (that is, the case of group symmetry for pure states is "equidiagonal"). The rank of operators $\hat{R}_\omega$ determines the multiplicity of representations $U_\omega (g)$ in the representation $U (g)$ in $\mathcal{U}$.

(5) *Suppose that the multiplicity of representations $U_\omega (g)$ in the representation $U (g)$ is unity. Then the operators $\hat{R}_\omega$ are one-dimensional, $\hat{R}_\omega = \psi_\omega^+ \otimes \psi_\omega$, and Eqs.* (2.3.3.16) *and* (2.3.3.17) *assume the form*

$$\Big(\int \hat{S}_\omega \, d\omega / c_\omega - \hat{I} \Big) \hat{\mu} = 0, \quad \Big(\sum \hat{S}_\omega \, d\omega / c_\omega - \hat{I}_\omega \Big) \hat{\mu} = 0,$$

where $\hat{S}_\omega = \mathrm{Tr}_{\mathcal{H}_\omega} \hat{R}_\omega$, and $c_\omega = [\mathrm{Tr}_{\mathcal{U}_0} (\hat{S}_\omega \hat{\mu})]^{1/2}$.

In particular, if all $\hat{S}_\omega$ are commutative, then operator $\hat{\mu}$ is a multiple of the proper projector $\hat{\pi}$ of operators $\hat{S}_\omega$, which corresponds to the eigenvalues λ_ω with maximal $\mu = \Big(\int \lambda_\omega^{1/2} \, d\omega \Big)^2$ $\Big($or $\mu = \Big(\sum \lambda_\omega^{1/2} \, d\omega \Big)^2 \Big)$: $\hat{\mu} = \mu \hat{\pi}$, $(\hat{S}_\omega - \lambda_\omega) \hat{\pi} = 0$.

2.3.3.5 Application to Fields with Indeterminate Phase and Group Symmetry

To apply the above results to the case of a decomposable G-homogeneous family of density operators $S_x = \bigoplus_n R_x^{(n)}$ it is sufficient to supply all the spaces and operators in (2.3.3.11)-(2.3.3.16) with an index n and then sum over n. In particular, if the representations $U^{(n)} (g)$ in $\mathcal{H}^{(n)}$ are nth tensor powers of the representation of $U (g)$ in $\mathcal{X}$, then the solution of the problem of optimal recognition of audio and optical fields is reduced to finding the irreducible representations $U_\omega (g)$ contained in $U^{(n)} (g)$. The operators $R_\omega^{(n)}$ determining (2.3.3.16) and (2.3.3.17) are

$$\hat{R}_\omega^{(n)} = \int \hat{r}^{(n)} (g) \, U_\omega (g) \, dg$$

where

$$\hat{r}^{(n)} (g) = \psi^{(n)+} U^{(n)} (g^{-1}) \, \psi^{(n)}.$$

For example, if the states S_x are Gaussian, the family of signals $\{\varphi_x, x \in X\}$ is G-homogeneous: $U (g) \, \varphi_{gx} = \varphi_x$, and the correlation noise operator L (or N) is G-invariant: $U (g) \, L U^+ (g) = L$, then the family of the $R^{(n)}$ operators is also G-homogeneous with respect

to the appropriate tensor powers $U^{(n)}(g)$ of the representation $U(g)$ in the subspace $\mathfrak{U}$ generated by vector $\varphi = \varphi_{x_0}$ for a certain $x_0 \in X$. In this manner we can find the exact solution to the following problems: resolution of several nonorthogonal partially coherent signals or fields that form a homogeneous family of permutations with respect to a certain group (symmetry gtoups $S(r)$ and their subgroups), estimation of the time lag of pulsed signals and the carrier frequency in quasiperiodic signals (cyclic Z groups), joint measurement of the duration and the frequency of a wave packet (the symplectic group), separate or joint measurement of momenta and position of quantum systems (and ensembles of such systems) with r degrees of freedom (the $Z(r)$ groups), detection of photon polarization and electron spin (the $SU(2)$ group), detection of complex signals and fields with equal intensities of rank r against a thermal background (the $SU(r)$ groups and their subgroups), and the like.

References

2.1. L.L. Myasnikov and E.N. Myasnikova, *Automatic Recognition of Sound Patterns* (Leningrad: Energiya, 1970) (in Russian).

2.2. N. Bohr, *Naturwissenschafter* **16**: 245-257 (1928); J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Berlin: Springer, 1932).

2.3. J. von Neumann, *Mathematical Foundation of Quantum Mechanics* (Princeton, N.J.: Princeton Univ. Press, 1955).

2.4. V.P. Belavkin, *Stochastics* No. 1: 315-345 (1975).

2.5. V.P.Belavkin, *Radiotehn. i Electron.* **20**, No. 6: 1177-1185 (1975).

2.6. V.P. Belavkin, *Radio Engrg. Electron. Phys.* **21**, No. 1: 78-86 (1976).

2.7. P.A. Bakut and S.S. Shehurov, *Problemy Peredači Informacii* **4**, No. 1: 77-82 (1968).

2.8. C.W. Helstrom, *J. Statist. Phys.* **1**: 231-252 (1969).

2.9. R.S. Kennedy, *M.I.T. Res. Lab. Electron. Quart. Progr. Rep.* **110**: 142-146 (July 15, 1973).

2.10. H.P. Yuen and M. Lax, *IEEE Trans. Inform. Theory* **IT-21**: 125-134 (March, 1975).

2.11. C.W. Helstrom, *Quantum Detection and Estimation Theory* (New York: Academic Press, 1976).

2.12. A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland, 1982).

2.13. C.W. Helstrom, *Information and Control* **10**: 254-291 (March, 1975).

2.14. C.W. Helstrom, *Phys. Lett.* **25A**: 101-102 (1967).

2.15. H.P. Yuen and M. Lax, *Proc. IEEE* **58**: 1770-1773 (1970).

2.16. B.A. Grishanin and R.L. Stratonovich, *Problemy Peredači Informacii* **6**, No. 3: 15-23 (1970).

2.17. V.P. Belavkin and B.A. Grishanin, *Problemy Peredači Informacii* **8**, No. 3: 103-109 (1972).

2.18. A.S. Holevo, in: *Proc. 2nd Japan-USSR Symp. Prob. Theo.*, vol. 1 (Kyoto, Japan, Aug. 1972): pp. 22-40.

2.19. V.P. Belavkin, *Radio Engrg. Electron Phys.* **17**: 2028-2032 (1972).

2.20. V.P. Belavkin, *Radio Engrg. Electron. Phys.* **17**, No. 12: 2032-2038 (1972).

2.21. V.P. Belavkin and B.A. Grishanin, *Problems of Information Translation* **9**, No. 3: 209-215 (1973).

2.22. R.L. Stratonovich, *Stochastics* No. 1: 87-126 (1973).

2.23. B.A. Grishanin, *Radio Engrg. Electron. Phys.* **18**, No. 4: 572-577 (1973).

2.24. A.S. Holevo (Kholevo), *J. Multivariate Anal.* No. 3: (1973).
2.25. V.P. Belavkin and R.L. Stratonovich, *Radio. Engrg. Electron. Phys.* 18, No. 9: 1349-1354 (1973).
2.26. V.P. Belavkin, *Problems Control Inform. Theory* 3: 47-62 (1974).
2.27. V.P. Belavkin and A.G. Vancjan, *Radio Engrg. Electron. Phys.* 19, No. 7: 1397-1401 (1974).
2.28. V.P. Belavkin, *Problems Control Inform. Theory* 4, No. 3: 241-257 (1975).
2.29. V.P. Belavkin, in: *Reports Delivered at the 6th Conference on Coding Theory and Information Transfer*, vol. 6, (Moscow-Tomsk: 1975): pp. 13-18 (in Russian).
2.30. V.P. Belavkin, *Zarubezhnaya Radio Elektronika* No. 5: 3-29 (1975).
2.31. V.P. Belavkin, *Teoret. Mat. Fiz.* No. 3: 213-222 (1976).
2.32. A.S. Holevo, in: *AMS Translation Proc. Steklov Math. Inst.* Issue 3 (1978)
2.33. V.P. Belavkin, *Problems Control Inform. Theory* 7, No. 5: 345-360 (1978).
2.34. V.P. Belavkin, *Radiotehn. i Elektron.* 25, No. 7: 1445-1453 (1980).
2.35. V.P. Belavkin, in: *Optimization Techniques* (Proc. of the 9th IFIP Conference on Optimization Techniques, Warsaw, September 4-8, 1979), Part I, K. Iracki, K. Malanowski, and S. Walukiewicz (eds.) (Berlin: Springer, 1980): pp. 141-149.
2.36. J.D. Gabor, *Proc. Reg. Phys. Soc. London, Sect. B* 64: 449 (1951).
2.37. H.J. Caulfield (ed.), *Handbook of Optical Holography* (New York: Academic Press, 1979): p. 230.
2.38. Yu.P. Pyt'ev, *Kibernetika*, No. 3: 126-134 (1973).
2.39. V.P. Maslov, *Théorie des perturbations et méthodes asymptotiques* (Paris: Gauthier-Villars, 1972).
2.40. A.A. Kuriksha, *Quantum Optics and Optical Location* (Moscow, Sov. radio, 1973) (in Russian).
2.41. C.W. Helstrom, *J. Opt. Soc. Amer.* 59: 164-175 (1969).
2.42. C.W. Helstrom, *J. Opt. Soc. Amer.* 60: 659-666 (1970).
2.43. R.J. Glauber, *Phys. Rev.* 130: 2529-2539 (1963).
2.44. J.R. Klauder and E.C.G. Sudarshan, *Fundamentals of Quantum Optics* (New York: W.A. Benjamin, 1968).
2.45. S.A. Akhmanov and A.S. Chirkin, *Statistical Phenomena in Nonlinear Optics* (Moscow: Moscow Univ. Press, 1971) (in Russian).
2.46. P. Halmos, *Summa Brasil. Math.* 2: 125-134 (1950).
2.47. M.A. Neumark, *C.R. Acad. Sci. d'URSS* 41: 359-361 (1943).
2.48. D.G. Luenberger, *Optimization by Vector Space Methods* (New York: Wiley, 1969).
2.49. H. Yuen and M. Lax, *IEEE Trans. Inform. Theory* IT-19, No. 6: 740-750 (1973).
2.50. C.R. Rao, *Linear Statistical Inference and Its Application* (New York: Wiley, 1965).

3 Mathematical Models in Computer-component Technology: Asymptotic Methods of Solution

V. G. Danilov, V. P. Maslov and *K. A. Volosov*

3.0 A Brief Survey

The most important electrophysical parameters of modern microelectronics devices and the quality, reliability, and effectiveness of the technology for designing such devices depend to a great extent on how extensively and intelligently the results of the modeling of the various stages in the technological processes are employed. There exists a vast literature devoted to this subject (e.g. see [3.1]).

Active employment of modeling techniques results in the following:

(a) designing of microelectronics devices takes less time;

(b) designing errors are eliminated;

(c) designing and manufacturing expenses are reduced.

Modeling in the field considered here is a way of optimizing the separate manufacturing processes, such as lithography, etching, and precipitation, that has gained wide acceptance and proved its worth. It is convenient for study of the complex links between the competitive physical mechanisms in successive multistage technological processes.

In designing computer components, modeling is employed to calculate the modes of functioning of these components. Its success has called for development of more exact models, which in most cases prove to be nonlinear.

The various stages of the production technology in microelectronics include the control of processes involved in the transfer of momentum, mass, electric charge, and energy. Although the physical processes manifesting themselves in these transfer mechanisms are highly diverse, they are described by equations of a special type, namely, the equations of momentum transfer (say, in the Navier-Stokes form) and quasilinear parabolic equations of diffusion, chemical kinetics, and energy of charge transfer.

At present in the study of transfer processes with a view to obtaining results that give a more precise description of physical reality, the focus has shifted to nonlinear mathematical models.

An important feature of computer-component production technology is that the equations constituting the mathematical models contain small parameters. This is due to the competition between

the various forces acting simultaneously in a model. Thus, a small parameter emerges when we apply the method of similarity and dimensional analysis [3.2] and is connected with scaling criteria.

The presence of a small parameter enables the researcher to apply asymptotic methods, say, those discussed in [3.3]. This, in turn, makes it possible to considerably broaden the class of problems that allow for analytical solutions and to consider the effect of the inhomogeneity of the medium and nonlinearity of the process. Asymptotic formulas make it possible to calculate the characteristics of a process in much less time and with considerably fewer other resources of a computer than by direct numerical integration. This aspect becomes especially important when flexible technological modules and complexes are introduced.

All mathematical models of transfer processes can be divided into two large classes. The first covers models in which the transfer coefficient is constant, that is, does not depend on the transferred quantity u. The second covers models in which the transfer coefficient $K(u)$ is a function of the transferred quantity u. In the second class the models of special interest are those in which the function $K(u)$ has a singularity at a certain value of the transferred quantity $u = \text{const}$, that is, a derivative of this function, $\partial^\alpha K/\partial u^\alpha$, $\alpha \geqslant 1$, has a discontinuity or, in other words, experiences a jump. From the mathematical point of view this second class has localized solutions and is characterized by a finite speed of propagation of perturbations.

The methods used in studying the second class have been discussed in detail in [3.3]. The present paper is restricted to examples of mathematical models that describe the diffusion of a light beam along an optical wave guide, the heat transfer in a superconductor, and the like.

The first class of models has long been developed in the literature. There are more than one hundred papers on the subject, but very few mathematical studies have employed nontrivial asymptotic methods. The present paper fills this gap to some extent. We will discuss these methods using models of processes of precipitation, oxidation, diffusion, thermal conduction, and the like.

The plan for the discussion follows. In Section 3.1 we explain the mathematical statements of the above-posed problems and provide examples of exact solutions. In these the solution of the initial problem is expressed in terms of solutions of certain nonlinear ordinary differential equations known as standard equations. In the simplest cases the solutions prove to be self-similar (or invariant).

In Section 3.2 we give the necessary data on the properties of, and methods used for studying, standard equations.

In Section 3.3 and in the subsequent sections we solve the problems posed in 3.1. The models employed describe the respective processes

with sufficient accuracy. For this reason the mathematical problems are quite complicated, and it is usually impossible to construct exact solutions. However, the presence of a small parameter makes it possible to use asymptotic methods for constructing the solution. And it has been found that the asymptotic solutions can be constructed from the solutions of the standard equations. In Section 3.3 we consider a time-dependent model of silicon oxidation in dry oxygen.[1] In Section 3.4 we discuss a model for silicon oxidation in halogen-containing media.

Section 3.5 is devoted to models of mass transfer. For one, we build an asymptotic solution of the precipitation problem. In Section 3.6 the topic is the diffusion of light in an active medium and the propagation of nonlinear thermal waves. Section 3.7 is devoted to models associated with Ginzburg-Landau equations. Finally, in Section 3.8 we give typical exact and asymptotically bounded as $\varepsilon \to 0$ solutions (which we term bounded synergets) of quasilinear and semilinear hyperbolic and parabolic equations that emerge in problems associated with microelectronics component-production technology.

3.1 Models of Stages of Production and the Functioning of Computer Components

3.1.1 Models of Stages of Production of Computer Components

In this section we formulate the problems that lead to quasilinear and semilinear parabolic equations and systems of such equations. It has been demonstrated that in dimensionless variables these equations contain a small parameter. In accordance with the plan outlined in the brief survey, this section gives the mathematical models for all the physical and chemical processes considered in the paper. The solutions to the problems formulated are given in subsequent chapters.

3.1.1.1 Processes of Silicon Oxidation

In modern production technology of microelectronics components, the heterogeneous processes, that is, processes that proceed in the bulk or at an interface between two or more phases, are employed in obtaining various film coatings, in surface and bulk alloying of semiconductors, in epitaxial growth of single-crystal materials, and in gettering of various harmful impurities in microelectronics components [3.4, 3.5].

---

[1] The most complete physical model of this process can be found in the works of N. A. Kolobov [3.3-3.6].

One of the most important basic processes in the production of microelectronics components is the high-temperature oxidation of silicon, which is widely used in manufacturing high-quality insulator-semiconductor systems (IS systems). Under thermal oxidation three phases participate in the formation of IS systems, namely, the gas (the oxidizer), the oxide film, and the surface region of the semiconductor substrate, and the oxidation reaction takes place at the interface between the two solid phases, which considerably complicates the picture of the processes occurring in the system and thus

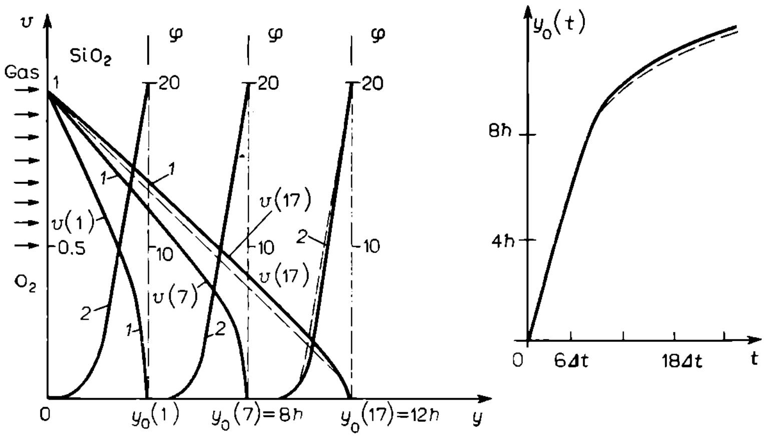


Fig. 3.1

requires developing a detailed physical model, whose perfection determines entirely the range of applicability and the accuracy of the mathematical models suggested.

The physical complexity of the oxidation process leads to an extremely complex general model [3.3]. The methods used to study this model can be applied to models of other heterogeneous processes.

Generally the oxidation of semiconductors amounts to the interaction of these semiconductors with oxidizing agents—oxygen, water, carbon and nitrogen oxides, and the like—leading to the formation of soluble, volatile, or stable oxides [3.4, 3.5].

The mathematical model of the oxidation process consists of an equation describing the diffusion of oxygen in silicon in the field of the space charge (Figure 3.1) and a Poisson equation describing the potential distribution. Curves *1* depict the distribution of oxygen concentration in silicon dioxide and curves *2*, the distribution of potential.

In dimensionless form, the diffusion equation is

$$\frac{\partial v}{\partial t} - \frac{1}{\mathrm{Pe}} \frac{\partial^2 v}{\partial y^2} + \frac{1}{\mathrm{Pe}} \frac{\partial}{\partial y}\left(v \frac{\partial \varphi}{\partial y}\right) = 0, \tag{3.1.1.1}$$

where y, t, x_0, t_0, $\mathrm{Pe} = x_0^2/(t_0 D)$, $\varphi = e\Phi/kT$, and $v = C(x, t)/C_0$ are, respectively, the dimensionless coordinate, the dimensionless time, the characteristic size (thickness of the silicon dioxide film), the characteristic time, the Péclet number (with D the diffusion coefficient), the dimensionless potential, and the dimensionless concentration. In the expression for the dimensionless potential, $\Phi(x, t)$ is the dimensional potential of the space charge at the interface between the silicon and silicon dioxide, k the Boltzmann constant, T the dimensional temperature, and e the electron charge. Finally, in the expression for the dimensionless concentration, C_0 is the dimensional concentration of oxygen at the gas-solid interface (the solid is the silicon dioxide film).

The experimental data [3.4, 3.5] suggest that $\mathrm{Pe} < 1$, which makes it possible to introduce a small parameter, $\varepsilon = \mathrm{Pe}$. The boundary conditions for the diffusion equation have the form

$$v(0, t, \varepsilon) = 1, \quad v(y_0(t), t, \varepsilon) = 0. \tag{3.1.1.2}$$

Here $y_0 = y_0(t)$ is the unknown dimensionless coordinate of the internal silicon-silicon dioxide interface; this will be found in the course of solving the problem.

To calculate the potential distribution in the oxide we must solve the appropriate Poisson equation, which in the one-dimensional case assumes the form

$$\frac{d^2\Phi(x, t)}{dx^2} = -\frac{|e|\, Q(x, t)}{\varepsilon\varepsilon_0}, \tag{3.1.1.3}$$

where $Q(x, t)$ is the space charge in the oxide, and ε and ε_0 are the dielectric constant of the oxide and the permittivity of empty space.

The physical model of the process implies (see [3.3-3.5]) that at the internal interface between the silicon and the silicon dioxide there is a negative potential Φ_0, whose size is determined by the energy balance of the free energy liberated in oxidation.

If the mobile charge carriers obey the Maxwell-Boltzmann statistics (and this agrees entirely with the case considered here), then

$$\begin{aligned} Q(x, t) &= |e|\, C(x)\left[\exp\left(-\frac{e\Phi(x, t)}{kT}\right) - \exp\left(\frac{e\Phi(x, t)}{kT}\right)\right], \\ \frac{d^2\Phi}{dx^2} &= \frac{2|e|\, C(x)}{\varepsilon\varepsilon_0} \sinh\left(\frac{e\Phi(x)}{kT}\right). \end{aligned} \tag{3.1.1.4}$$

Diffusion processes proceed at a considerably lower rate than electrodynamical processes. Hence, in the model considered the

electric potential at the silicon-silicon dioxide interface depends weakly on time.

In dimensionless form, the equation for the potential has the form

$$\frac{d^2\varphi}{dy^2} = v(y, t, \varepsilon)\left(\frac{x_0}{L_D}\right)^2 \sinh\varphi. \tag{3.1.1.5}$$

Let us assume that $\varepsilon^{3/2} = L_D/x_0 < 1$, which agrees with the case considered, with $L_D^2 = \varepsilon\varepsilon_0 kT/2\,|e|^2 C_0$ the square of the Debye shielding length.

The boundary conditions have the form

$$\varphi(0, t, \varepsilon) = 0, \quad \varphi(y_0(t), t) = \Phi_0 < 0. \tag{3.1.1.6}$$

Remark 3.1.1.1 Generally speaking, the equation for potential Φ must be solved in the region $0 < x < \infty$ with the boundary condition $\Phi(-\infty, t) = 0$. We can assume that the potential is shielded by mobile charge carriers of both signs (negatively charged molecular ions of oxygen, O_2^-, and oxygen vacancies of the oxide, V_0^+, may serve as such carriers). Then, estimating the potential via Gauss's theorem, we can formulate the problem for φ on a finite interval. The boundary conditions then have the form of (3.1.1.6) and are approximate. Such an approach makes it possible to disregard the processes occurring outside the oxide film.

Summing up, we arrive at the following system of equations:

$$\begin{aligned} &\varepsilon^3 \frac{\partial v}{\partial t} - \varepsilon^2 \frac{\partial^2 v}{\partial y^2} + \varepsilon^2 \frac{\partial}{\partial y}\left(v \frac{\partial\varphi}{\partial y}\right) = 0, \\ &\varepsilon^3 \frac{\partial^2\varphi}{\partial y^2} = v(y, t)\sinh\varphi. \end{aligned} \tag{3.1.1.7}$$

The movement of the silicon-silicon dioxide interface is determined by the flux balance equation

$$\frac{dy_0}{dt} = C^*\left(-\frac{\partial v}{\partial y} + v\frac{\partial\varphi}{\partial y}\right)\Big|_{y=y_0(t)}, \tag{3.1.1.8}$$

where C^* is a given constant [3.3-3.5].

Thus, Eqs. (3.1.1.7) and (3.1.1.8) together with the boundary conditions (3.1.1.2) and (3.1.1.6) constitute a complete mathematical model of the process of thermal oxidation of silicon when the kinetics of thermal oxidation is close to steady-state.

3.1.1.2 Distribution of Chlorine in Silicon Dioxide

Let us examine the diffusion of chlorine in a silicon crystal. When gaseous chlorine comes into contact with a silicon plate, there occurs classical diffusion of chlorine in silicon. The distribution of the chlorine concentration in the neighborhood of

point $x = 0$ is shown in Figure 3.2. This distribution resembles a boundary layer distribution, and on the whole the chlorine is dissolved poorly in silicon. The situation changes drastically when oxygen appears at the beginning of the solid-phase chemical reaction. This reaction leads to the emergence of an elastic stress field at the silicon-silicon dioxide interface, and the field promotes an anomalous distribution of chlorine as an impurity. Let us study more carefully the physical reasons for this anomalous distribution of chlorine in the Si-SiO_2 system.

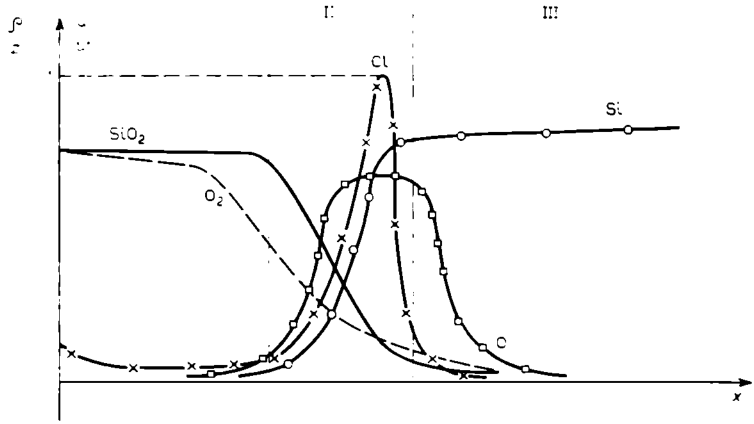


Fig. 3.2

It has been established experimentally that at the Si-SiO_2 boundary, which moves at a rate v_{oxide}, the distribution of the impurity has a pronounced maximum whose position is closely related to the transition region between the silicon and the silicon dioxide [3.4-3.6]. Formation of the transition region depends to a great extent on the change in the free volume of the unit cell of the crystal when the transition is made from the Si phase to the SiO_2 phase.

We suggest the following mechanism of formation of anomalous distribution of impurities [3.3]. The impurity particles diffuse into the solid phase and simultaneously there occurs a solid-phase chemical reaction, which leads to a local deformation at the Si-SiO_2 interface and, in turn, generates an elastic stress field (note that in an elastic stress field the rate of solid-phase reaction increases [3.7]). The products of the reaction of chlorine with the substances of the transition region are assumed to be stable only in the elastic stress field and disintegrate outside this region.

Figure 3.2 shows the experimental distribution ([3.4], and later [3.8]) of chlorine concentration in the Si-SiO_2 system. The maximal value of the concentration is taken for unity. In the same figure we depict the assumed distribution of the elastic stress field. In the bulk of SiO_2, the high gradient of the chlorine concentration is in opposition to the flux and is maintained owing to the elastic field stress.

An elastic stress field appears not only in chemical reactions accompanied by a significant change in the volume of the crystal lattice, as is the case in the above example, but may also be generated by

(a) a gradient in the impurity concentration,

(b) a temperature gradient produced in pulsed heating of the sample,

(c) electro- and magneto-elastic effects,

(d) ultrasound waves.

On the basis of these assumptions we offer in this section a phenomenological mathematical model suggested by N.A. Kolobov together with the present authors that takes into account the chemical reaction and the dependence of the diffusion coefficient on the impurity concentration and its gradient [3.4].

The existence of solutions that can describe, say, the distribution of an admixture of chlorine was discovered in mathematical models of transfer processes [3.3].

Let $C\,(x,\,t)$ be the concentration of chlorine in the solid, with x and t the coordinate and time, and let the diffusion coefficient have the form

$$D=\frac{D_0}{1+\beta_1\,|\,\partial C/\partial x\,|}\,, \tag{3.1.1.9}$$

where β_1 and D_0 are constants. This formula describes the decrease in the diffusion coefficient as the concentration gradient grows, with D_0 the diffusion coefficient at a given temperature. By virtue of formula (3.1.1.9), which is a generalization of Fick's law, the diffusion flux is expressed thus:

$$J=\frac{D_0}{1+\beta_1\,|\,\partial C/\partial x\,|}\,\frac{\partial C}{\partial x}\,.$$

We introduce the function $F_0\,(C)$, which models the solid-phase chemical reaction proceeding in the medium:

$$F_0(C)=C^q\bar{G}(C),\quad \bar{G}(a_1)=0,\quad \bar{G}(a_2)=0,\quad \bar{G}(a_3)=0,$$
$$\bar{G}(0)=0,\quad \partial\bar{G}/\partial C\,\Big|_{\substack{C=a_1\\ \bar{G}(C)>0}}<0,\quad d\bar{G}/dC\,|_{C=a_3}<0, \tag{3.1.1.10}$$

with $\bar{G}\,(C)>0$ for $0<C<a_1$, $\bar{G}\,(C)>0$ for $a_2<C<a_3$, and $C_{\max}=a_3$, $q>1$.

An example of function $\bar{G}\,(C)$ may be the following function: $\bar{G}\,(C)=AC\,(C-a_1)\,(a_2-C)\,(C-a_3)\exp C$, with $A=\text{const}\gg 1$.

The equation describing the diffusion of the impurity has the form

$$\frac{\partial C}{\partial t'}-\frac{\partial}{\partial x}\left(\frac{D_0}{1+\beta_1\,|\,\partial C/\partial x\,|}\,\frac{\partial C}{\partial x}\right)-F_0(C)=0. \qquad (3.1.1.11)$$

Now let us go over to dimensionless variables in this equation, assuming that

$$C(x,\,t)/C_{\max}=v(x,\,t),\quad x/x_0=y,\quad t'/t_0=t,\quad x_0^2/t_0D_0=\mathrm{Pe}.$$

with $C_{\max}$, x_0, and t_0 the maximal concentration and the characteristic size and time, respectively, and putting $a_i/C_{\max}=\bar{a}_i$. Substitut-

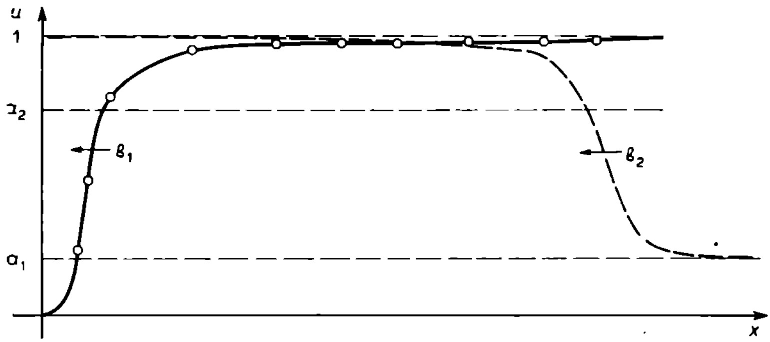


Fig. 3.3

ing into the expression for the Péclet number the values of the quantities characterizing the given process, namely,

$$x_0\sim 3\times 10^{-4}\ \mathrm{m},\quad t_0\sim 10^3\ \mathrm{s},\quad D_0\sim 10^{-11}\ \mathrm{m^2/s},$$

we obtain the estimate

$$\mathrm{Pe}\sim 10.$$

Assuming that Pe^{-1} is a small parameter, $\varepsilon=\mathrm{Pe}^{-1}$, we can rewrite Eq. (3.1.1.11) in the new variables thus:

$$\frac{\partial v}{\partial t}-\varepsilon\frac{\partial}{\partial y}\left(\frac{1}{1+\beta_2\,|\,\partial v/\partial y\,|}\,\frac{\partial v}{\partial y}\right)-\frac{1}{\varepsilon}F_1(v)=0, \qquad (3.1.1.12)$$

with β_2 a positive constant and function $F_1(v)$ satisfying the following conditions:

$$F_1(v)=v^qG(v),\quad G(\bar{a}_i)=0,\quad i=1,\,2,\quad G(1)=0,$$
$$G(0)=0,\quad \partial G/\partial v\,|_{v=\bar{a}_1}<0,\quad \partial G/\partial v\,|_{v=1}<0,$$

where $G(v)$ is positive for $v\in(0,\,\bar{a}_1)$ and negative for $v\in(\bar{a}_1,\,\bar{a}_2)$.

The boundary conditions for Eq. (3.1.1.12) assume the form

$$v\,|_{x\to-\infty}=0, \quad v\,|_{x\to+\infty}=a_1.$$

We will study this equation in Section 3.4.

The solution to Eq. (3.1.1.12) is depicted in Figure 3.3 and describes a formed impurity front moving into the bulk of the crystal. The rate of motion of this front is bounded above:

$$b<2\sqrt{\frac{1}{\mathrm{Pe}}\left|\frac{d}{dv}\left(\frac{G(v)\,v^q}{1+\beta_2\,|\,dv/dt\,|}\right)\right|_{v=\bar{a}_2}}.$$

The authors of this paper developed, together with N.A. Kolobov, a general model for the process of anomalous diffusion of chlorine:

$$\frac{\partial u}{\partial t}-\frac{1}{\mathrm{Pe}_1}\frac{\partial^2 u}{\partial x^2}+\gamma_1 u+\gamma_2\rho=0,$$

$$\frac{\partial v}{\partial t}-\frac{1}{\mathrm{Pe}_2}\frac{\partial^2 v}{\partial x^2}+\delta_1 v+\delta_2\rho+\delta_3 z=0,$$

$$\frac{\partial w}{\partial t}-\beta_1 v-\beta_2 u-\beta_3\rho+\beta_4 z=0,$$

$$\frac{\partial \rho}{\partial t}-\frac{1}{\mathrm{Pe}_3}\frac{\partial^2 \rho}{\partial x^2}+l_1 v+l_2 v=0,$$

$$\frac{\partial z}{\partial t}-\frac{1}{\mathrm{Pe}_4}\frac{\partial}{\partial x}\left(k\left(\frac{\partial z}{\partial x}\right)\frac{\partial z}{\partial x}\right)+\mu_1 v+\mu_2 u+\mu_3 z=0.$$

Here u, v, w, ρ, and z are the concentrations of, respectively, molecular oxygen, silicon, silicon dioxide, atomic oxygen, and the chlorine impurity (see Figure 3.2). The region where the given system is studied can roughly be divided into three zones: the silicon dioxide zone (I), the transition region between the silicon and the silicon dioxide with an admixture of chlorine (the region is filled with $\mathrm{Si}_\alpha\mathrm{O}_\beta\mathrm{Cl}_\gamma$ of unknown stoichiometry and is characterized by strong inner stresses caused by the difference in the lattice parameters) (II), and the silicon phase zone (III). By Pe_i we denote the diffusion Péclet numbers $\mathrm{Pe}_i=L^2/(D_i t_0)$, $i=1, 2, 3, 4$, where L and t_0 are the characteristic thickness of the silicon dioxide layer and the characteristic time during which the process sets in (i.e. the time it takes for the law of growth of the silicon dioxide layer with time to become linear [3.3, 3.6]), D_i are the diffusion coefficient:

$$D_i=\begin{cases} D_{i1} & \text{in zone I,}\\ D_{i2} & \text{in zone II,}\\ D_{i3} & \text{in zone III,}\end{cases}$$

and γ_i, δ_i, β_i, l_i and μ_i are the reaction rates fixed for each zone. The coefficient of chlorine diffusion in the elastic stress field is

inversely proportional to the concentration gradient:

$$K\left(\frac{\partial z}{\partial x}\right)=\frac{1}{1+A\mid \partial z/\partial x\mid^{\beta}}.$$

The boundary conditions have the form

$$u\,|_{x=0}=1,\quad v\,|_{x=0}=0,\quad w\,|_{x=0}=1,\quad \rho\,|_{x=0}=0,$$
$$z\,|_{x=0}=0,\quad u\,|_{x\to\infty}=0,\quad v\,|_{x\to\infty}=1,\quad w\,|_{x\to 0}=0,$$
$$\rho|_{x\to\infty}=0,\quad z\,|_{x\to\infty}=0.$$

The constants μ_1, μ_2, μ_3, δ_3, and β_4 change sign depending on the sign of $\partial z/\partial x$, namely, they are all positive for $\partial z/\partial x$ negative and negative for $\partial z/\partial x$ positive. Further discussion of this model lies outside the scope of the present article.

Hence, there exists a real possibility in a number of cases of selecting a chemical reaction that will enable alloying various crystals with poorly soluble impurities. This, in turn, may lead to fabrication of new microelectronics elements with unique properties.

3.1.1.3 Sorption, Adsorption, and Precipitation

These processes occur in the gaseous and liquid phases. The physical bases of oxidation adsorption at the interface between the gas and the hydroxylated surface of the oxide are discussed in great detail by N.A. Kolobov in the Appendix to [3.3] (see also [3.4, 3.5, 3.9]). The Appendix also contains the model of the process:

$$\frac{\partial u}{\partial t}+\mu v\frac{\partial u}{\partial y}+\frac{1}{m_1}\frac{\partial w}{\partial t}=\varepsilon\frac{\partial}{\partial y}\left(\frac{\partial u}{\partial y}\right),\ \frac{\partial w}{\partial t}=f(u,\ w),\quad (3.1.1.13)$$

with μ and m_1 constants, u the dimensionless concentration of the substance contained in the medium surrounding the surface, w the dimensionless concentration of the substance deposited on the surface, y the dimensionless coordinate in the direction normal to the surface, f the isotherm of the process, and $\varepsilon = Dt_0/L^2$ the small parameter (here D is the diffusion coefficient, t_0 the characteristic time, and L the characteristic size). Similar models describe various modifications of the precipitation process, say, chemisorption.

In the liquid phase the precipitation is sometimes accompanied by chemical reactions, with the result that the mathematical model assumes the form

$$\frac{\partial u}{\partial t}+\mu v\frac{\partial u}{\partial y}+\frac{1}{m_1}\frac{\partial w}{\partial t}=\frac{1}{\mathrm{Pe}}\frac{\partial}{\partial y}\left(\frac{\partial u}{\partial y}\right)+F(u),$$
$$\frac{\partial w}{\partial t}=f(u,\ w),\qquad (3.1.1.14)$$

with μ and m_1 constants. The first equation describes mass transfer accompanied by chemical reactions, and the second (known as the

kinetic equation) describes the properties of the chemical process (its kinetics). In (3.1.1.14) the functions u and w stand for the concentration of the substance contained in the phase surrounding the surface and the concentration of the substance on the surface, $v(x, t)$ is the rate of admission of the substance, $F(u)$ is a smooth function describing the variation in u due to chemical reactions, f is the isotherm of the process, and Pe is the diffusion Péclet number, $\mathrm{Pe} = L^2/(Dt_0)$, which is a small or large parameter. Methods for constructing asymptotic solutions to (3.1.1.14) are developed in [3.3].

3.1.1.4 Microwelding of Current Leads in Computer Components

Microwelding is widely used in the manufacture of non-detachable contacts in assembling computer components and other electronic devices. To calculate the heat condition in this process it is usually necessary to know the temperature field generated by a number of point heat sources combined in a certain configuration. The results are then employed to select a configuration of the sources, a shape of heat pulses, and a design so that one of the following conditions is met:

(a) no complex phase transitions have time to occur in the microwelding zone, or

(b) the thermal effect must be such that only selected phase transitions occur in the microwelding zone in the necessary direction.

In the first case we must ensure that the temperature in the vicinity of the heat source drops off rapidly in time; in the second we must select the given temperature distribution.

The process of heat transfer in microwelding is described by a heat equation, which in dimensionless form is

$$\rho(u, x) C(u, x) \frac{\partial u}{\partial t} - \varepsilon^2 \langle \nabla, \lambda(u, x) \nabla u \rangle + R(x, t) = 0, \tag{3.1.1.15}$$

where $\langle , \rangle$ stands for the scalar product, ρ, c, and λ are the density, specific heat, and heat conductivity and constitute dimensionless positive functions (these piecewise linear functions are replaced in a mathematical model with smooth functions), $R(x, t)$ is the dimensionless source function describing the configuration of the heat sources ($x \in R^2$, $t \in R_+$), $u = T/T_0$ is the dimensionless temperature, and $\varepsilon^2 = \lambda_0 t_0/(\rho_0 C_0 L^2)$ is a parameter, with ρ_0, C_0, λ_0, L, and t_0 the characteristic values of density, specific heat, thermal conductivity, distance, and time, respectively.

The problem involving only one electrode has been considered in [3.10]. If there are n electrodes, the function $R(x, t)$, which models

the distribution of the sources, has the form

$$R(x, t) = \sum_{i=1}^{n} A_i(x, t) \exp\{-a_i^{-1} \sum_{j=1}^{3} (x - x_j)^2\}, \quad (3.1.1.16)$$

where the A_i are finite smooth functions, x_j the dimensionless coordinates of the electrodes, and a_i constants that determine the fractional width of the heat pulses.

3.1.1.5 Liquid Epitaxy

This method is employed in the manufacture of microelectronics elements so as to obtain high-quality-low-defect single-crystal semiconductor materials. The epitaxy method [3.11] consists of crystallizing a substance from its solution or melt (the solution of the melted components that we wish to crystallize in a low-melting solvent). The parameters and structure of the epitaxial layer are controlled on Earth by varying the temperature, the rate of growth of the layer, the cooling rate, container geometry, the placing of the substrate, etc. In outer space in conditions of low acceleration there is the possibility of varying the size of the vector of residual accelerations that directly influence the quality of the layer.

The epitaxy process is described by a system of Navier-Stokes equations in the Boussinesq approximation of slow flow of a fluid [3.11]:

$$\omega = \nabla^2\psi, \quad \frac{\partial \omega}{\partial t} + u\frac{\partial \omega}{\partial x} + v\frac{\partial \omega}{\partial y} = \nabla^2\omega + F,$$

$$\frac{\partial C}{\partial t} - u\frac{\partial C}{\partial x} + v\frac{\partial C}{\partial y} = \frac{\nabla^2 C}{\mathrm{Sc}},$$

$$u = \frac{\partial \psi}{\partial y}, \quad v = -\frac{\partial \psi}{\partial x},$$

where $F = \left(\frac{\partial C}{\partial x}\cos\varphi + \frac{\partial C}{\partial y}\sin\varphi\right)\mathrm{Gr}$ are the bulk forces, $\mathrm{Gr} = g\beta L^3 C_0/\nu^2$ the Grashof number, $\beta = (\partial\rho/\partial C)_{\rho, T}\rho^{-1}$, $0 < y < 1$, $0 < x < 1$. Here g, β, L, C_0, ν, ρ, u, and v the acceleration generated by the mass force, the expansibility at constant pressure and temperature, the characteristic distance, the concentration calculated from the composition-property diagram, the kinematic viscosity, the density, and the components of the velocity vector; $\mathrm{Sc} = \nu/D$ is the Schmidt number, and ω, ψ, and C are the dimensionless hydrodynamic vortex and stream functions and the concentration. The direction of the acceleration caused by the mass force makes an angle φ with the y axis.

The boundary conditions are

$$C = C_D, \; u = 0, \; v = 0, \; x = 0, \; x = 1, \; 0 < y < 1,$$

$$\partial C/\partial y = 0, \; u = 0, \; v = 0, \; y = 0, \; y = 1, \; 0 < x < 1,$$

where C_D is a constant.

The Grashof number may be either greater or less than unity, which makes it possible to introduce a small parameter, ε. The Schmidt number exceeds unity considerably, say, when GaAs layers are grown.

In the present model we assume that the temperature of the melt is constant and that only the flow of the liquid and the concentration distribution of the crystallizing component in the low-melting solvent are considered. The liquid-solid phase transition is accompanied by liberation of latent heat of fusion, and the temperature distribution also affects the process.

3.1.1.6 A Mathematical Model for Phase Transitions

In dimensionless variables this model has the form [3.12]:

$$\begin{gathered}
\varepsilon \frac{\partial u}{\partial t} - \frac{\partial}{\partial x} K(u) \frac{\partial u}{\partial x} - \gamma\left(u^n - u_0^n(x, t)\right) = 0, \\
\beta\left(\frac{dx_1(t)}{dt}\right) = \left[K(u) \frac{\partial u}{\partial x}\right]_{x=x_{1-}}^{x=x_{1+}}, \quad u(x_1, t) = u_{n1}, \qquad (3.1.1.17) \\
\lim_{x \to \pm\infty} \partial\left(u^n - u_1^n(x, t)\right)/\partial x = 0, \qquad n = 1 \text{ or } n = 4,
\end{gathered}$$

where γ is a constant, $u_1(x, t)$ a given function describing the temperature in the external medium, u_{n1} the temperature of the phase transition (a known constant), $\varepsilon = \mathrm{Pe} < 1$ is one small parameter, $\beta = \sigma/(\lambda T_0^3) < 1$ is the second small parameter of the problem, and σ, λ, and T_0 are the normalized Stefan-Boltzmann constant, which allows for the degree of blackness of the body, the characteristic value of thermal conductivity, and the temperature. The thermal flux in the phase transition is specified by the following relationship:

$$K(u) \frac{\partial u}{\partial x} = \beta\left(u_{n1}^n - u^n\right).$$

Here we have assumed that heat transfer by radiation plays an important role.

3.1.1.7 Electron and Hole Transfer in Semiconductors

The mathematical model of this process [3.13] consists of a system of parabolic and elliptic equations, which in dimensionless variables have the form

$$\begin{gathered}
\nabla^2 \varphi = Q(n - p - f), \\
\frac{\partial n}{\partial t} = \langle \nabla, \mu_n (\nabla n - n \nabla \varphi) \rangle + g, \qquad (3.1.1.18) \\
\frac{\partial p}{\partial t} = \langle \nabla, \mu_p (\nabla p + p \nabla \varphi) \rangle - g.
\end{gathered}$$

where φ is the electric potential, n and p are the electron and hole concentrations, n_1 and p_1 are the maximal values of these concentrations, q, μ_p, and μ_n are the charge of the particles and their mobilities, g is the recombination generation function, $Q = q_{n_0}(\varepsilon_1\varepsilon_0\varphi_T)$,

$$\mu_l = \left(\mu_{\min}^l + \frac{\mu_{\max}^l}{1+(f/f_{0i})^{\lambda_i}}\right)\frac{1}{1+|\nabla\varphi/E_{0l}|^{1/\beta_l}},$$

$g = (np - n_0^2)[\tau_p(n + n_1) + \tau_n(p + p_1)]$, n_0 is the intrinsic electron concentration in the semiconductor, ε_1 and ε_0 are the dielectric constant of the medium and the permittivity of empty space, $f(x, y, t)$ is a known smooth function describing the total density of impurities in the semiconductor (donors and acceptors), the variable l assumes one of two values (n or p), the subscript i assumes two values (1 and 2), τ_p and τ_n are the lifetimes of holes and electrons prior to recombination, and φ_T, q, $\mu_{\max}^l$, $\mu_{\min}^l$, E_{0l}, β_l, and λ_i are fixed constants.

The quantity $\varepsilon \overset{\text{def}}{=} Q$ ($Q \sim 10^{-3}$) is the small parameter in this model. For a concentration characteristic of the problem discussed, a second small parameter (the diffusion Péclet number) can be constructed using the characteristic values of mobility μ_{0l}, the characteristic time interval (say, τ_p), and the characteristic distance.

3.1.2 Models of Functioning of Computer Components

In this section we describe models for heat transfer in a superconducting matrix, the diffusion of a light beam in an optical fiber, and the formation of stable spin waves in ferromagnetic liquid.

3.1.2.1 Heat Conduction in a Superconducting Matrix

In view of the recent discovery of a new type of superconducting materials with a transition temperature of about 90-100 K and higher, the use of such materials in technology and computer components becomes ever more important. The design of a superconductor resistance matrix presupposes good heat conduction and a good thermal contact of the superconducting material with the nitrogen thermostat. Superconducting materials with a high critical field (type II superconductors) usually have a low mean free path of electrons and, hence, a lower thermal conductivity, which potentially makes them unstable [3.14, 3.15]. The copper layers, which form the framework of the matrix, are insulators (compared with the superconductor) and possess high thermal conductivity.

Let us consider a superconducting resistance matrix immersed in a nitrogen thermostat with a temperature T_0 [3.15, 3.16]. In the destruction of the superconducting state (the phase transition from

the superconducting state to the normal conducting state), the temperature of the superconductor changes from T_0 to the critical temperature T_c above which the superconducting state is destroyed. There exists a distinct boundary between the superconducting state and the normal metal state (Figure 3.4). Three zones can be specified on the diagram: ω_1, or the region of normal conductivity, ω_2, or the

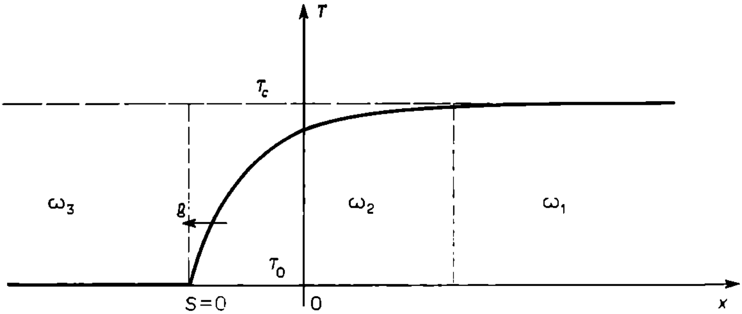


Fig. 3.4

narrow transition region (which is natural for type II superconductors), and ω_3, or the region of the superconducting state. The boundary between ω_2 and ω_3 can move in either direction. Since the phase transition occurs fairly rapidly, we must retain the second time derivative in the heat equation [3.14-3.17].

Heat conduction in a composite superconductor consisting of the resistance matrix and the superconducting layers is described in a uniform magnetic field by the following equation [3.16-3.17]:

$$\rho C(u)\left(\frac{\partial u}{\partial t}+t_0\frac{\partial^2 u}{\partial t^2}\right)-\varepsilon\frac{\partial}{\partial x}\left(K(u)\frac{\partial u}{\partial x}\right)+F(u,x,t)=0, \qquad (3.1.2.1)$$

where ρ is the dimensionless density of the medium, $K(u)$ is a positive function describing the thermal conductivity coefficient of the medium, $u = T/T_0$, $F = [-J_0^2R + \alpha(u-1)]_\mu$, $\mu = \text{const}$, $\alpha = T_0/(T_c - T_0)$, $J_0 = J - J_m(1 - B/B_m - \alpha(u-1))$, $1 - B/B_m - \alpha(u-1) > 0$, with J the current flowing in the resistor matrix, J_m and B_m are the upper values of the critical current and of the magnetic field induction for a given superconductor, T_0 is the temperature maintained in the nitrogen thermostat, $R = a_1 + a_2B$ is the resistance of the composite semiconductor, $C = a_3u^3$ is the volume specific heat of the superconductor, $a_i = \text{const} > 0$, $i = 1, 2, 3$, $F(u, x, t)$ is the source-sink distribution function (the rate of heat release), and $\varepsilon^{-1} = \rho_0C_0L^2/(\lambda_0t_0)$ is the

small parameter of the problem, with the quantities supplied with the "0" subscript being the characteristic values of the density, specific heat, thermal conductivity, and time, and L being the characteristic distance in the problem.

3.1.2.2 Diffusion of a Light Beam in an Optical Fiber and in an Optically Active Medium

Let us consider the problem of light diffusion in a medium with small transparent reflecting particles. At a certain moment in time the medium is illuminated by light from a source. The source is then switched off. Obviously, the intensity of the beam at the wave guide exit will become equal to zero, since the beam undergoes multiple reflection in the medium, that is, the beam must travel a very long path before it arrives at the exit. Employing probabilistic ideas to the motion of a particle and calculating the reflection and absorption probabilities of the particle, the authors of [3.18] arrive at the following hyperbolic equation:

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} + \frac{2\nu}{c^2}\frac{\partial u}{\partial t},$$

where c and ν are the speed of light in the medium and the frequency.

When light passes through an active medium, we must add, starting from a certain value of light intensity, a certain number of particles to the intensity of the flux. The flux intensity reaches its maximum and then decreases as the population inversion of the upper energy levels decreases due to the presence of excited particles. In this case the intensity is described by the equation

$$\frac{\nu}{c^2}\frac{\partial u}{\partial t} + \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) - F(u) = 0, \tag{3.1.2.2}$$

$F(a_0) = 0$, $F(a_1) = 0$, $dF/du\,|_{u=a_i} \neq 0$, $i = 0, 1$. In dimensionless variables, Eq. (3.1.2.2) has in one of the approximations the form

$$\varepsilon\frac{\partial \tilde{u}}{\partial \tilde{t}} + \varepsilon^2\frac{\partial^2 \tilde{u}}{\partial \tilde{t}^2} - \varepsilon^2\frac{\partial^2 \tilde{u}}{\partial \tilde{x}^2} - \tilde{F}(\tilde{u}) = 0, \tag{3.1.2.3}$$

where the variables marked with a tilde placed over them are dimensionless.

3.1.2.3 Formation of Stable Spin Waves in a Ferromagnetic Substance

This process can be used in the memories of modern computers and in the biological computers of the future [3.19, 3.20]. The basic trend in the development of modern memory devices is

still an increase in the storage density and storage space. We believe that this trend will continue.

For certain values of the parameters there appears in a ferromagnetic film a steady-state (stable) modulation in the form of a quasi-periodic "focusing." This is reflected in the distribution of the maxima and minima of the magnetization vector. As the magnitude of the "supercriticality" increases as the system goes through a series of stages of "development," against the background of large modulations there appear modulations on a smaller scale. The generation of capillary waves on the surface of a liquid may serve as an analog of this physical process and is easily observed in experiments [3.21-3.24]. The mathematical models of these phenomena are extremely close.

The same phenomenon is observed in the generation of Langmuir waves in a plasma and the generation of waves on the surface of a liquid insulator in an electric field or of a liquid ferromagnetic substance placed in a variable magnetic field. The generation of steady-state waves was also observed on the surface of melted metal heated by modulated ionic beams [3.23].

All the phenomena mentioned above are described by a single mathematical model.

To study the dynamics of the wave field we can employ averaged equations. The deviation of the level of the surface, $\zeta(x, y, t)$, can be represented in the form of a sum of four terms [3.23]:

$$\zeta = \frac{1}{2}\left[a_+ e^{ikx} + a_- e^{-ikx} + b_+ e^{iky} + b_- e^{-iky}\right] e^{-i\omega t} + \text{c.c.},$$

where c.c. stands for the complex conjugate of the first term, k is the wave number, and ω the frequency.

The common method of deriving the truncated equations is to employ the results of the Hamiltonian description of the nonlinear interaction of gravitation-capillary waves [3.24]. When going over to the wave packet approximation, we must retain four packets corresponding to two pairs of counterrunning waves.

The system of equations for the envelopes has the form [3.24]

$$\begin{aligned}
&\frac{\partial a_\pm}{\partial t} \pm v_g \frac{\partial a_+}{\partial x} - \frac{i}{4}\frac{v_g}{k}\frac{\partial^2 a_\pm}{\partial x^2} - \frac{i v_g}{2k}\frac{\partial^2 a_\pm}{\partial y^2} + \gamma a_\pm \\
&\quad = i(H + F b_+ b_-) a_\mp^* + i a_\pm \left[T |a_\pm|^2 - S |a_\mp|^2 - R(|b_+|^2 + |b_-|^2)\right], \\
&\frac{\partial b_\pm}{\partial t} \pm v_g \frac{\partial b_+}{\partial y} - \frac{i}{4}\frac{v_g}{k}\frac{\partial^2 b_\mp}{\partial y^2} - \frac{i}{2}\frac{v_g}{k}\frac{\partial^2 b_\mp}{\partial x^2} \\
&\quad + \gamma b_\pm = i(H + F a_+ a_-) b_\mp^* + i b_\pm \left[T |b_\pm|^2 - S |b_\mp|^2 - R(|a_+|^2 + |a_-|^2)\right],
\end{aligned} \tag{3.1.2.4}$$

where T, S, R, F, and H are constants, γ is the damping constant, and v_g is the group velocity. This model will be analyzed in Section 3.7.

3.1.3 Examples of Simple Mathematical Models

In this section we discuss simple models referring to equilibrium precipitation dynamics and a model of modulation of steady-state spin waves in a ferromagnetic substance. We also build new solutions to the Ginzburg-Landau equation. Some of the mathematical models discussed in Sections 3.1.1 and 3.1.2 allow exact solution in particular cases; for others no exact solutions are known. We will give the exact solutions in cases where they are known and thereby demonstrate a class of solutions that constitute bounded (as $\varepsilon \to 0$) synergets. These will be discussed in Sections 3.5 and 3.7.

3.1.3.1 The Exact Solution to the Model of Equilibrium Precipitation Dynamics

When the influx of the precipitated substance to the surface is constant, the mathematical model of the dynamics of the molecular process is [3.9]

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial y} + \frac{\varepsilon}{m}\frac{\partial \varphi}{\partial t} = \varepsilon D\frac{\partial^2 v}{\partial y^2}, \quad v = \left(\frac{b}{a-\varphi}\right)^{\lambda},$$

$$v|_{y\to-\infty} \to 1_-, \quad \partial v/\partial y|_{y\to-\infty} \to 0, \quad v|_{y=Ct} = (b/a)^{\lambda},$$

$$-\infty \leqslant y \leqslant Ct.$$

(Here C is an unknown constant that must be found, and a and b are constants that satisfy the condition $a - b = 1$.) Assuming that $v = v(\tau)$, $\varphi = \varphi(\tau)$, and $\tau = (y - Ct)/\varepsilon$ (in this example we assume that $\varepsilon = 1$, so that $\tau = y - Ct$), we arrive at an ordinary differential equation for the function $v(\tau)$:

$$-C\frac{dv}{d\tau} + u\frac{dv}{d\tau} - \frac{C}{m}\frac{d}{d\tau}(a - b/v^{1/\lambda}) = D\frac{d^2v}{d\tau^2}. \tag{3.1.3.1}$$

The boundary conditions for $v(\tau)$ are

$$v(-\infty) = 1, \quad \frac{dv}{d\tau}(-\infty) = 0, \quad v(0) = (b/a)^{\lambda}. \tag{3.1.3.2}$$

Integrating (3.1.3.1), we get a first-order equation:

$$(-C + u)v - \frac{C}{m}(a - b/v^{1/\lambda}) = D\frac{dv}{d\tau}. \tag{3.1.3.3}$$

Assuming that $v \to 1$ and $dv/d\tau \to 0$ as $\tau \to -\infty$ and allowing for the condition $a - b = 1$, we find that $C = um/(m-1)$. A parti-

cular solution to Eq. (3.1.3.3) has the form

$$\tau=\frac{m+1}{u}D\int\limits_{(b/a)^{\lambda}}^{v}\frac{dv}{v-\varphi}=\frac{m+1}{u}D\int\limits_{(b/a)^{\lambda}}^{v}\frac{v^{1/\lambda}\,dv}{v^{(\lambda+1)/\lambda}-av^{1/\lambda}+b}. \tag{3.1.3.4}$$

The denominator of the integrand vanishes at $v = 1$, since $a - b = 1$. If we denote expression on the left-hand side of (3.1.3.4) by $I_1(v)$, we get $\tau = I_1(v)$.

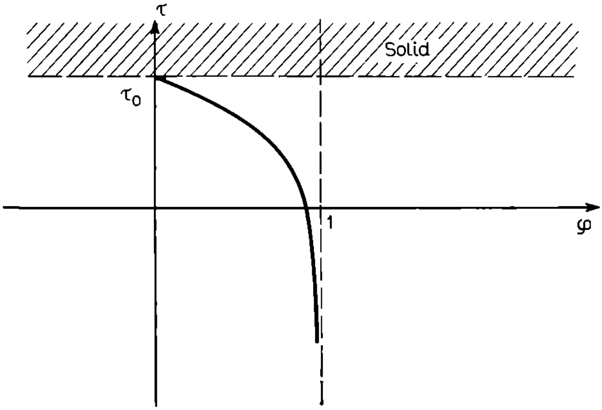


Fig. 3.5

Generally speaking, the improper integral in $I_1(v)$ is divergent at $v = 1_-$, that is, $\tau \to -\infty$ as $v \to 1_-$, but the inverse function $v = v(\tau)$ exists because the conditions of the theorem on inverse functions are met.

Let us restrict our discussion to the case where $\lambda = 1$. We have

$$\begin{aligned}I_1(v) &= D\frac{m+1}{u}\int\limits_{b/a}^{v}\frac{v\,dv}{v^2-av+b}\\ &= D\frac{m+1}{u}\Big[\ln|v^2-av+b|^{1/2}\\ &+\frac{a}{2}\int\frac{dv}{(v-1)(v-b)}\Big]\Big|_{b/a}^{v}\\ &= D\frac{m+1}{u}\ln\left[|v^2-av+b|\left|\frac{v-1}{v-b}\right|^{a/(1-b)}\right]^{1/2}\Bigg|_{b/a}^{v}.\end{aligned} \tag{3.1.3.5}$$

The solution $\tau(v)$ is depicted in Figure 3.5, and the curve

$$\varphi(\tau)=\begin{cases} a-b/v^{1/\lambda} & \text{if} \quad v>(b/a)^{\lambda}, \\ 0 & \text{if} \quad v\leqslant(b/a)^{\lambda} \end{cases}$$

is depicted in Figure 3.6.

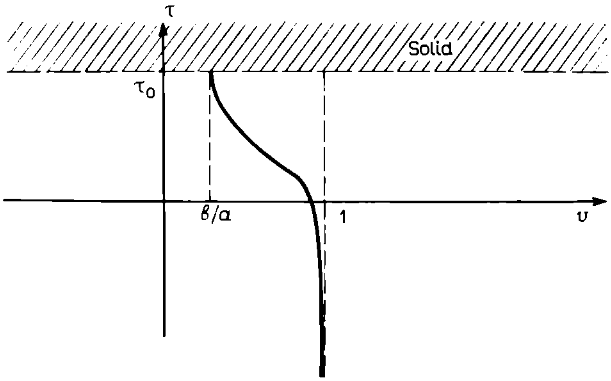


Fig. 3.6

The case involving a variable flow velocity $u(y, t)$ and other cases of models of precipitation will be discussed in greater detail in Section 3.5.

3.1.3.2 Spin Waves in a Ferromagnetic Substance

Let us take the simple example of the modulation of steady-state spin waves in a ferromagnetic substance (the statement of problems of this class has been carried out in Section 3.1.2). This phenomenon can easily be elucidated if we employ the model of capillary waves on the surface of a liquid.

Let us examine the nonlinear interaction of two wave packets. If in Eqs. (3.1.2.4) we put $a = a(y, t)$ and $b_{\pm} = 0$ for the amplitude of the envelope, we obtain the following equation:

$$\frac{\partial a}{\partial t}-\frac{i}{2}\frac{v_g}{k}\frac{\partial^2 a}{\partial y^2}+\gamma a=iHa^{*}+ia\left[T+S\right]\,|\,a\,|^{2}+i\beta a, \tag{3.1.3.6}$$

where $H = Gk/(4\omega)$, $G = \text{const}$, $T = 0.0625\,\omega k^2$, $S = 0.625\omega k^2$, $\gamma = 2\nu k$ is the damping constant, and $\beta = \text{const}$.

In this situation the authors of [3.23] observed a series of grooves on silicon (Figure 3.7) at the following values of the constants:

density $\rho = 0.9 \times 10^3$ kg/m^3, surface tension $\alpha = 23 \times 10^{-3}$ N/m, kinematic viscosity $\nu = 4 \times 10^{-6}$ m^2/s, wavelength $\lambda = 3.2 \times 10^{-3}$ m, $\omega = 2\pi \times 70$ rad/s, and the characteristic size of the region studied was 0.2 m. At these values of the physical parameters the "deep water" approximation is valid.

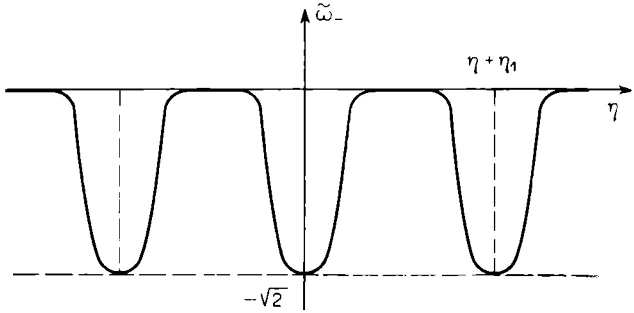


Fig. 3.7

The typical modulation period is equal to $m\lambda$, with m varying between 2 and 4. The characteristic ratios in this case are $\gamma/\omega = 0.07$ and $v_g/\gamma^\lambda = 3.4$.

These values of the parameters suggest that Eq. (3.1.3.6) can be made dimensionless and that there is a small parameter, $0 < \varepsilon < 1$, in the model.

Suppose that $\eta = y/(\lambda m)$, $\tau = t\gamma$, $u = a\,[(S + T)/\gamma]^{1/2}$, $\sigma = \beta/\gamma$, $h = H/\gamma$, and $\varepsilon^2 = v_g/(2k\lambda^2\gamma\sigma m^2) < 1$, where ε is the small parameter. Then Eq. (3.1.3.6) assumes the form

$$\frac{\partial u}{\partial \tau} - i\varepsilon^2 \frac{\partial^2 u}{\partial \eta^2} + u = ihu^* + iu\,|\,u\,|^2 - i\sigma u. \qquad (3.1.3.7)$$

The value $h = 1$ is said to be the threshold value, and the first stable "quasiperiodic focusing" mode is observed at $h = 1 + \varepsilon h_0$, where εh_0 is the small supercriticality term.

We will seek the solution to Eq. (3.1.3.7) in the form

$$u\,(\tau, \eta, \varepsilon) = U\,(\eta, \varepsilon) + \varepsilon W\,(\tau, \eta, \varepsilon) + O\,(\varepsilon^2).$$

Then for the functions U and W we arrive at the following equations:

$$-i\varepsilon^2 \frac{\partial^2 U}{\partial \eta^2} + iU\,|\,U\,|^2 - iU\sigma + iU^* - U = 0, \qquad (3.1.3.8)$$

$$\frac{\partial W}{\partial \tau} - i\varepsilon^2 \frac{\partial^2 W}{\partial \eta^2} + W = ih_0 U^* + iW\,|\,U\,|^2$$
$$+ 2i\,|\,UW\,|\,W - i\sigma W. \qquad (3.1.3.9)$$

If we put
$$U = n_1 + in_2,$$
we get the following system of equations:
$$\begin{aligned} &\varepsilon^2 \frac{\partial^2 n_2}{\partial \eta^2} + (-1+\sigma) n_2 - n_1 - n_2 (n_1^2 + n_2^2) = 0, \\ &-\varepsilon^2 \frac{\partial^2 n_1}{\partial \eta^2} + (1-\sigma) n_1 - n_2 + n_1 (n_1^2 + n_2^2) = 0. \end{aligned} \tag{3.1.3.10}$$

Near the threshold of parametric wave excitation [3.23], it is necessary that $n_1 = n_2 C = n$, where the constant C in view of

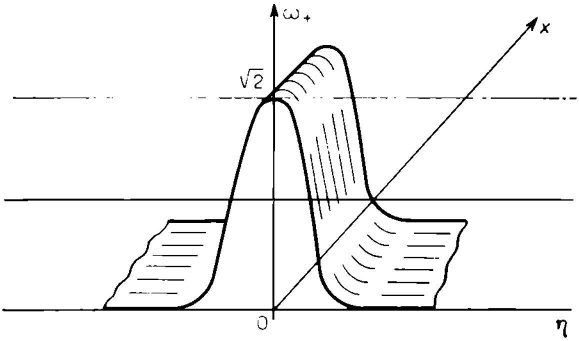


Fig. 3.8

(3.1.3.10), identically unity, $C \equiv 1$ (see also [3.21, 3.22]). Then from system (3.1.3.10) we obtain the equation
$$-\varepsilon^2 \frac{d^2 n}{d\eta^2} + \sigma n - 2n^3 = 0, \tag{3.1.3.11}$$
whose exact solution has the form [2]
$$n = (\sigma/2)^{1/2} \omega (\eta \sqrt{\sigma}/\varepsilon), \quad \omega_\pm = \pm \sqrt{2}/\cosh \xi, \ \ \xi = \eta \sqrt{\sigma}/\varepsilon.$$

Equation (3.1.3.11) has two steady-state stable solutions: $\omega = 1$ and $\omega = -1$. The function $|\omega_-|$ is exponentially close to unity at a sufficient distance from point $\eta = 0$, namely, the following estimates hold true:
$$\omega_- = \begin{cases} \varepsilon^N + o(\varepsilon^N) & \text{if } N > 1 \text{ and } \eta > -N\varepsilon \ln \varepsilon / \sqrt{\sigma} = \delta, \\ \varepsilon^N + o(\varepsilon^N) & \text{if } \eta < N\varepsilon \ln \varepsilon / \sqrt{\sigma}. \end{cases} \tag{3.1.3.12}$$
Similar estimates hold true for ω_+.

[2] The fact that parameter ε is small will be employed below to construct the asymptotic solution (3.1.3.17). The exact solution is well-known and can be obtained in a trivial manner.

The function $\omega_-(\xi)$ can be matched with the steady-state solutions via the so-called decomposition-of-unity procedure. We denote this new function by $\tilde{\omega}_-(\xi)$, with $\xi = \eta\sqrt{\sigma}/\varepsilon$.

Let

$$
\begin{aligned}
E_1(\xi) &= \begin{cases} 0 & \text{if} \quad \eta < \delta, \\ 1 & \text{if} \quad \eta > \delta + \delta_1, \quad \delta_1 > 0, \end{cases} \\
E_2(\xi) &= \begin{cases} 0 & \text{if} \quad \eta > -\delta, \\ 1 & \text{if} \quad \eta < -\delta - \delta_1 \end{cases}
\end{aligned}
\tag{3.1.3.13}
$$

be two smooth, infinitely differentiable functions. Then

$$
\tilde{\omega}_-(\xi) = \begin{cases} [\omega_-(\xi)(1 - E_1(\xi))] & \text{if} \quad \xi \geqslant 0, \\ [\omega_-(\xi)(1 - E_2(\xi))] & \text{if} \quad \xi \leqslant 0. \end{cases}
\tag{3.1.3.14}
$$

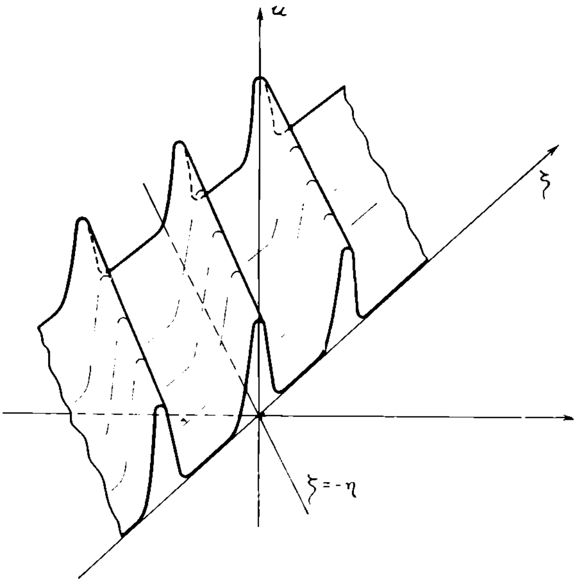


Fig. 3.9

Let us describe how the function $\tilde{\omega}_-$ can be used to construct a single groove (Figures 3.8 and 3.9). Note that Eq. (3.1.3.11) and the functions ω_- are invariant with respect to the translation group,

The functions E_i are also shifted by a quantity η_1:

$$E_1\left(\frac{\eta+\eta_1}{\varepsilon}\right)=\begin{cases}0 & \text{if} \quad \eta+\eta_1<\delta,\\ 1 & \text{if} \quad \eta+\eta_1>\delta+\delta_1, \quad \delta>0,\end{cases}$$
$$E_2\left(\frac{\eta+\eta_1}{\varepsilon}\right)=\begin{cases}0 & \text{if} \quad \eta+\eta_1>-\delta,\\ 1 & \text{if} \quad \eta+\eta_1<-\delta-\delta_1.\end{cases} \tag{3.1.3.15}$$

In a manner similar to that employed in (3.1.3.14), we can set up the function $\tilde{\omega}_-(\xi)$:

$$\tilde{\omega}_-(\xi+\xi_1)$$
$$=\begin{cases}[\omega_-(\xi+\xi_1)(1-E_1(\xi+\xi_1)) & \text{if} \quad \xi+\xi_1>0,\\ [\omega_-(\xi+\xi_1)(1+E_2(\xi+\xi_1)) & \text{if} \quad \xi+\xi_1<0.\end{cases}$$
$$\xi_1=\text{const}>\delta_1. \tag{3.1.3.16}$$

Then functions (3.1.3.14) and (3.1.3.16) can be employed to construct a single groove (the asymptotic solution to Eq. (3.1.3.7)):

$$u=(1+i)(\sigma/2)^{1/2}\tilde{\omega}_-(\xi)+O(\varepsilon). \tag{3.1.3.17}$$

A more detailed study of problems of this kind will be carried out in Section 3.7. In the two-dimensional case the similar solution to the equation $\varepsilon^2\left(\frac{\partial^2\omega}{\partial\eta^2}+\frac{\partial^2\omega}{\partial\zeta^2}\right)-\omega+\omega^3=0$ has the form $\omega=\sqrt{2}/\cosh\left(\frac{\eta+\zeta}{\varepsilon\sqrt{2}}\right)$.

3.2 Properties of Standard Equations

Here we list the properties of ordinary differential equations that will be needed in our future discourse.

3.2.1 Semilinear Differential Equations

In this section we discuss the basic properties of semilinear standard equations.

The exact solution to the model of diffusion of a light beam in an optical fiber (the model constructed in Section 3.1.2) has the form

$$u(x, t, a)=\chi\left(\frac{x+bt}{\varepsilon}\right),$$

where the function $\chi(\tau)$ satisfies the standard differential equation

$$b\frac{d\chi}{d\tau}-\frac{d}{d\tau}\left((\mu(\chi)-b^2)\frac{d\chi}{d\tau}\right)-R(\chi)=0. \tag{3.2.1.1}$$

This equation belongs to the class of ordinary differential equations of the type

$$b\frac{d\chi}{d\tau}-\frac{d}{d\tau}\left(K(\chi)\frac{d\chi}{d\tau}\right)-R(\chi)=0, \tag{3.2.1.2}$$

which represents all the standard equations encountered in this paper.

At $K(\chi)=1$ such equations are known as semilinear:

$$b\frac{d\Theta}{d\xi}-\frac{d^2\Theta}{d\xi^2}-R(\Theta)=0. \tag{3.2.1.3}$$

The properties of monotone semilinear solutions to Eq. (3.2.1.3) for an $R(\Theta)\not\equiv 0$ that has only two roots on the closed interval [0, 1] have been investigated in two situations [3.25-3.29]:

$$\begin{gathered} R\in C^1([0,1]),\quad R(0)=0,\quad R(1)=0,\\ \left.\frac{dR}{d\Theta}\right|_{\Theta=0}\times\left.\frac{dR}{d\Theta}\right|_{\Theta=1}<0; \end{gathered} \tag{3.2.1.4}$$

$$\begin{gathered} R\in C^1([0\ \ 1]),\ R(0)=0,\ R(1)=0,\\ \left.\frac{dR}{d\Theta}\right|_{\Theta=0}=0,\ \left.\frac{dR}{d\Theta}\right|_{\Theta=1}<0,\ R(\Theta)>0. \end{gathered} \tag{3.2.1.5}$$

Equation (3.2.1.3) in which the function $R(\Theta)$ satisfies conditions (3.2.1.4) is known as a Kolmogorov-Petrovskii-Piskunov (KPP) equation after the mathematicians who obtained these equations and studied them in the now classical work [3.26] and who arrived at nontrivial results. It is now well-known that when conditions (3.2.1.4) are met, Eq. (3.2.1.3) possesses smooth monotone solutions $0\leqslant\Theta(\xi)\leqslant 1$ only if $b\geqslant 2\sqrt{|dR/d\Theta|_{\Theta=0}|}$, with

$$\begin{gathered} \Theta(\xi)\sim 1-\exp(-l_0\xi)\ \text{as}\ \xi\to\infty,\\ \Theta(\xi)\sim\exp(l\xi)\ \text{as}\ \xi\to-\infty,\ b>2\sqrt{|dR/d\Theta|_{\Theta=0}|},\\ \Theta(\xi)\sim c_1\exp(l\xi)+c_2\xi\exp(l\xi)\ \text{as}\ \xi\to-\infty,\\ b=2\sqrt{|dR/d\Theta|_{\Theta=0}|}, \end{gathered} \tag{3.2.1.6}$$

when $dR/d\Theta\,|_{\Theta=0}>0$. When $dR/d\Theta\,|_{\Theta=0}<0$, one should simply substitute $1-\Theta$ for Θ. See [3.43] and the Remark on p. 320.

Equation (3.2.1.3) in which the function $R(\Theta)$ satisfies conditions (3.2.1.5) is known as the Zeldovich equation [3.3, 3.28] and differs in principle from the KPP equation. First, the solution to the Zeldovich equation that satisfies the conditions $\Theta(\xi)\to 1$ as $\xi\to\infty$ and $\Theta(\xi)\to 0$ as $\xi\to-\infty$ exists only for a single value of parameter b and behaves, as $|\xi|\to\infty$, in the same manner as (3.2.1.6). Second, a reversal of sign of $R(\Theta)$ in the Zeldovich equation leads to an equation with entirely different properties.

3.2.1.1 Kolmogorov-Petrovskii-Piskunov Equations

A particular case of Eq. (3.2.1.3) is the equation

$$b\frac{d\Theta}{d\xi}-\frac{d^2\Theta}{d\xi^2}-\Theta(1-\Theta)=0. \qquad (3.2.1.7)$$

This equation has an infinitude of solutions of the traveling-wave type, that is, functions that satisfy the conditions

$$0\leqslant\Theta\leqslant 1,\quad \Theta(-\infty)=0,\quad \Theta(+\infty)=1,$$
$$d\Theta/d\xi\neq 0 \text{ for } \tau\in(-\infty,\infty), \qquad (3.2.1.8)$$

with $b\geqslant b_{\min}=2$. The following theorem (see [3.26] and also [3.25]) holds true:

Theorem 3.2.1.1 *For every $b\geqslant 2$ there exists a solution to Eq.* (3.2.1.7) *satisfying conditions* (3.2.1.8).

The results of this theorem are generalized in [3.3] to incorporate equations of the (3.2.1.2) type in which the function $\Theta(1-\Theta)$ is replaced with an $R(\Theta)$ that is smooth for $0\leqslant\Theta\leqslant 1$ and satisfies the following conditions:

$$R(\Theta)>0 \text{ for } 0<\Theta<1,\ R(0)=R(1)=0,$$
$$\left.\frac{dR}{d\Theta}\right|_{\Theta=0}>0,\ \left.\frac{dR}{d\Theta}\right|_{\Theta=1}<0. \qquad (3.2.1.9)$$

As is well known, solutions of the traveling-wave type (i.e. solutions satisfying (3.2.1.8)) exist in this case if

$$b\geqslant b_{\min}=2\sqrt{|\,dR/d\Theta\,|_{\Theta=0}\,|}. \qquad (3.2.1.10)$$

Let us calculate the constants l and l_0 in (3.2.1.6). After we differentiate Eq. (3.2.1.3) with respect to ξ and apply L'Hospital's rule, we find that, as $\xi\to-\infty$,

$$b=l+|\,dR/d\Theta\,|_{\Theta=0}\,|/l. \qquad (3.2.1.11)$$

This function is depicted in Figure 3.10. Obviously, if $b>b_{\min}$ (a given constant), then

$$0<|\,l\,|<\infty,\ l=b/2-\sqrt{b^2/4-dR/d\Theta\,|_{\Theta=0}}\,.$$

Let us calculate the constant l_0. Following the same line of reasoning as in deriving (3.2.1.11), we find that, as $\xi\to\infty$,

$$b=l_0+\frac{|\,dR/d\Theta\,|_{\Theta=1}|}{l_0},\quad l_0\overset{\text{def}}{=}\left.\frac{d^2\Theta/d\xi^2}{d\Theta/d\xi}\right|_{\xi=\infty}<0. \qquad (3.2.1.12)$$

This function is depicted in Figure 3.11. Since $\Theta \to 1$ as $\xi \to \infty$ and the value of b (l) is finite (nonzero), we obtain

$$\lim_{\xi \to \infty} \left[-\frac{d^2\Theta/d\xi^2}{d\Theta/d\xi} \right] = l_0,$$
$$l_0 = -b/2 + \sqrt{b^2/4 + |dR/d\Theta|_{\Theta=1}|}.$$

Whence the following relationships:

$$\Theta = 1 - \exp(-|l_0|\xi) + o(\exp(-|l_0|\xi)), \quad \xi \to \infty,$$
$$d\Theta/d\xi = |l_0| \exp(-|l_0|\xi) + o(\exp(-|l_0|\xi))$$
$$= -|l_0|(1 - \exp(-|l_0|\xi)) + |l_0|$$
$$+ o(\exp(-|l_0|\xi))$$
$$= -|l_0|\Theta + |l_0| + o(\exp(-|l_0|\xi))$$
$$= |l_0|(1-\Theta) + o(\exp(-|l_0|\xi)), \quad \xi \to \infty.$$

The above reasoning proves the validity of Theorem 3.2.1.1.

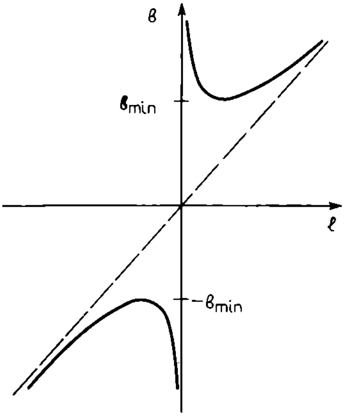


Fig. 3.10

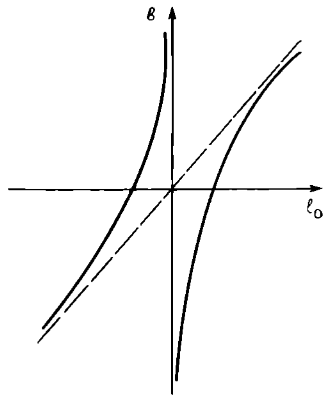


Fig. 3.11

In applications an important equation is the one involving the sink function:

$$b\frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} + R(\Theta) = 0, \quad R(\Theta) > 0 \text{ for } \Theta \in (0, 1).$$

This is combined with conditions (3.2.1.4) and the conditions

$$\Theta|_{\xi \to +\infty} = 1, \quad \Theta|_{\xi \to -\infty} = 0.$$

The velocity of a simple wave in this problem is negative,

$$b = -\left(\int_0^1 R(\Theta)\,d\Theta\right)\left(\int_{-\infty}^{+\infty}\left(\frac{d\Theta}{d\xi}\right)^2 d\xi\right)^{-1},$$

and the following estimates hold true:

$$\Theta \approx \exp(l\xi),\ \ l = b/2 + \sqrt{b^2/4 + dR/d\Theta\,|_{\Theta=1}} > 0,\qquad \xi \to -\infty,$$

$$\Theta \approx 1 - \exp(-l_1\xi),\ \ l_1 = -b/2 + \sqrt{b^2/4 - |dR/d\Theta\,|_{\Theta=1}|},\qquad \xi \to +\infty,$$

so that the wave travels with the velocity

$$b \leqslant b_{\min} = -2\sqrt{|\,dR/d\Theta\,|_{\Theta=1}\,|}.$$

Finally, here is an example that illustrates how important it is that the derivative of the source-sink function, $dR/d\Theta$, for $\Theta \in [0, 1]$ in Eq. (3.2.1.3) be continuous. In the papers [3.26, 3.27] this condition is ignored.

Let us consider the function

$$\Theta(\xi) = \begin{cases} \exp(-1/\xi) & \text{if } \xi \geqslant 0, \\ 0 & \text{if } \xi < 0, \end{cases} \tag{3.2.1.13}$$

which is smooth for $\xi \in R^1$ and satisfies the simple wave equation for equations of the KPP type [3.26]

$$b\frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} - \Theta\ln^2\Theta\,[b - \ln\Theta\,(2 + \ln\Theta)] = 0,\ \ \Theta \in [0, 1]. \tag{3.2.1.14}$$

In this case the function R in Eq. (3.2.1.3) has the form

$$R = \Theta\ln^2\Theta\,[b - \ln\Theta\,(2 + \ln\Theta)]$$

and has an unbounded first derivative at point $\Theta = 0$. This results in the localization of solution (3.2.1.13) to Eq. (3.2.1.14).

3.2.1.2 Zeldovich Equations

A particular case of Eq. (3.2.1.3) in which the function $R(\Theta)$ satisfies the conditions (3.2.1.5) is the following one:

$$b_0\frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} - \Theta^2(1 - \Theta) = 0. \tag{3.2.1.15}$$

The exact monotone solution to this equation has the form

$$\Theta(\xi) = \frac{1}{1 + \exp(-\xi/\sqrt{2})}. \tag{3.2.1.16}$$

The constant b_0, known as the Zeldovich constant, is in the present case equal to $1/\sqrt{2}$. As $|\xi| \to \infty$, we obtain the following estimates:

$$\Theta = \begin{cases} O(\exp(b_0\xi)) \text{ as } \xi \to -\infty, \\ 1-\exp(-b_0\xi)+o(\exp(-b_0\xi)) \text{ as } \to +\infty. \end{cases}$$

These results can be generalized to incorporate equations of the (3.2.1.3) type in which the function $\Theta^2(1-\Theta)$ is replaced with a smooth function $R(\Theta)$ satisfying conditions (3.2.1.5). As is known (see [3.3]), in this case the traveling-wave solution (i.e. satisfying condition (3.2.1.8)) exists only for a single value of constant b and, as $|\xi| \to \infty$, we arrive at estimates of the (3.2.1.6) type, with $l = l_0$.

The Zeldovich equations have been studied in [3.30] in connection with problems of the theory of combustion. The physics of these problems implies that the function $R(\Theta)$ must be nonnegative. However, a study of localized solutions to nonlinear equations (see Section 3.8 and the book [3.3]) leads to the need to study equations of simple waves for the Zeldovich equations with a nonpositive $R(\Theta)$.

A particular case of the Zeldovich equation with a nonpositive $R(\Theta)$ is

$$b\frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} + \Theta^2(1-\Theta) = 0. \tag{3.2.1.17}$$

Let us prove that the solution to this equation satisfying the conditions

$$d\Theta/d\xi \neq 0, \quad \Theta|_{\xi\to-\infty} \to 0, \quad \Theta|_{\xi\to+\infty} \to 1, \tag{3.2.1.18}$$

has no exponential asymptotic as $\xi \to -\infty$. The existence of such a solution is proved below.

Let us assume that the contrary is true, that is, that there are positive constants $l_1 < 0$ and $l_2 > 0$ such that

$$\Theta = \begin{cases} 1-\exp(l_1\xi)+o(\exp(l_1\xi)) \text{ as } \xi \to +\infty, \\ O(\exp(l_2\xi)) \text{ as } \xi \to -\infty. \end{cases}$$

Then, as $|\xi| \to \infty$, (3.2.1.17) yields

$$-bl_1 + l_1^2 + 1 = 0, \quad bl_2 - l_2^2 = 0.$$

Real solutions l_1 exist only if $b \geqslant 2$. The second equation implies that $b = l_2$. Note that the following equalities are valid:

$$b\int_{-\infty}^{\infty}\left(\frac{d\Theta}{d\xi}\right)^2 d\xi = -\int_0^1 \Theta^2(1-\Theta)\,d\Theta = -\frac{1}{12}.$$

Hence, b and l_2 are negative, which contradicts the initial assumption.

It can easily be verified that the solution to Eq. (3.2.1.17) satisfying the conditions (3.2.1.18) has the following asymptotic behavior:

$$\Theta(\xi) = O(1/|\xi|) \text{ as } \xi \to -\infty,$$
$$\Theta(\xi) = 1 - \exp(l_1\xi) + o(\exp(l_1\xi)) \text{ as } \xi \to +\infty, \quad (3.2.1.19)$$
$$l_1 = b/2 - \sqrt{b^2/4 - 1}.$$

The condition $b \geqslant 2$ is the necessary condition for the existence of a solution.

Suppose that the function $R(\Theta)$ in Eq. (3.2.1.3) is differentiable on the closed interval [0, 1] and satisfies the conditions

$$R(\Theta) < 0 \text{ for } 0 < \Theta < 1, \; R(0) = R(1) = 0, \qquad (3.2.1.20)$$
$$\left.\frac{dR}{d\Theta}\right|_{\Theta=0} = 0, \; \left.\frac{dR}{d\Theta}\right|_{\Theta=1} > 0.$$

Then the following assertion is true:

Lemma 3.2.1.1 *If conditions* (3.2.1.20) *are met, then there exists a solution to Eq.* (3.2.1.3) *satisfying conditions* (3.2.1.18).

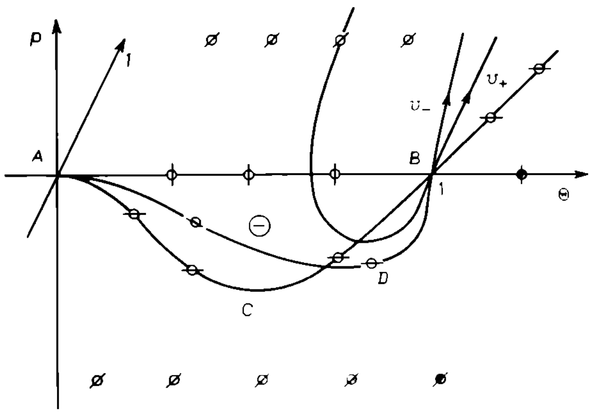


Fig. 3.12

Proof. The proof follows from an analysis of the phase portrait of the first-order ordinary differential equation

$$\frac{dp}{d\Theta} = \frac{bp - R(\Theta)}{p}, \quad b < 0,$$

to which Eq. (3.2.1.3) can be reduced (see Figure 3.12, where the curve ACB corresponds to the function $p = R(\Theta)/b$ and the "minus" sign singles out the region where the derivative $dp/d\Theta$ is negative).

The singular point A is a complicated singular point of the saddle-node type with a single distinct direction (direction 1 in Figure 3.12). The singular point B is a node for $|b| > 2\sqrt{dR/d\Theta|_{\Theta=1}}$ (a confluent node for $|b| = 2\sqrt{dR/d\Theta|_{\Theta=1}}$). The eigenvalues and eigenvectors of the linearized equation have the form

$$\lambda_{\pm} = b/2 \pm \sqrt{b^2/4 - dR/d\Theta|_{\Theta=1}}, \quad v_- = (\lambda_-, 1),$$
$$v_+ = (\lambda_+, 1).$$

The integral curve ADB, which corresponds to the solution to Eq. (3.2.1.17) that satisfies conditions (3.2.1.18), enters the singular point B along the direction of the eigenvector v_-. The proof of Lemma 3.2.1.1 is complete.

There exists a natural generalization of Lemma 3.2.1.1, namely,

Lemma 3.2.1.2 *Suppose that the function* $R(\Theta)$ *in Eq.* (3.2.1.3) *satisfies conditions* (3.2.1.20) *and point* $\Theta = 0$ *is an algebraic branch point of this function,* $R(\Theta) \sim \Theta^{\beta}$, $\beta > 1$. *Then the solution* $\Theta(\xi)$ *satisfying conditions* (3.2.1.18) *has the following asymptotic behavior*:

$$\Theta(\xi) = O(|\xi|^{-1/(\beta-1)}) \ \text{as} \ \xi \to -\infty,$$
$$\Theta(\xi) = 1 - \exp(l\xi) + o(\exp(l\xi)) \ \text{as} \ \xi \to +\infty,$$
$$l = b/2 - \sqrt{b^2/4 - dR/d\Theta|_{\Theta=1}} < 0.$$

Proof. The proof can be obtained in the same manner as was done with Lemma 3.2.1.1 and is not discussed here.

3.2.2 Nonlinear Standard Equations

In this section we consider a method for studying nonlinear standard equations. The method is based on the reduction of the standard equation to known equations for simple waves for the KPP equation [3.26] of the Zeldovich equations [3.30]. The properties of the solutions to these equations are given below. The main results concerning the properties of localized solutions of standard equations are contained in Theorems 3.2.2.1 and 3.2.2.2.

Let us take an ordinary differential equation with constant coefficients,

$$b\frac{d\chi}{d\tau} - \frac{d}{d\tau}\left(K(\chi)\frac{d\chi}{d\tau}\right) - F(\chi) = 0, \tag{3.2.2.1}$$

where $K(\chi)$ and $F(\chi)$ are positive for $\chi \in (0, 1)$, $K(0) = 0$, $K(1) > 0$, and $K(\chi) \sim \chi^{k-1}$ as $\chi \to 0$, $k > 1$. By $\Theta(\xi)$ we denote the smooth monotone solution to

$$b\frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} - R(\Theta) = 0, \tag{3.2.2.2}$$

where $dR(\Theta)/d\Theta$ is a continuous function of $\Theta \in [0, 1]$.

Theorem 3.2.2.1 *The transformation*

$$K(\chi)\frac{d\chi}{d\tau}=\frac{d\Theta}{d\xi}(\tau(\chi)) \qquad (3.2.2.3)$$

reduces Eq. (3.2.2.2) *to Eq.* (3.2.2.1), *where* $F(\chi)=R(\chi)/K(\chi)$.

Remark 3.2.2.1 The inverse of function $\chi(\tau)$, or $\tau(\chi)$, exists because of the assumption that $\Theta(\xi)$ is monotone. Indeed, Eq. (3.2.2.3) and the monotonicity of $\Theta(\xi)$ imply the monotonicity of $\chi(\tau)$.

Remark. 3.2.2.1′ We get the same results using $\theta=\exp[(b+\sqrt{b^2-b_{\min}^2})\tau/2]$, $\tau\to-\infty$ (see p. 320).

Proof. Transformation (3.2.2.3) leads to the following:

$$\frac{d\chi}{d\tau}=\frac{1}{K(\chi)}\frac{d\Theta(\tau(\chi))}{d\xi},$$

$$\frac{d}{d\tau}\left(K(\chi)\frac{d\chi}{d\tau}\right)=\frac{d^2\Theta}{d\xi^2}\frac{1}{d\Theta(\tau(\chi))/d\xi}\frac{d\chi}{d\tau}$$

$$=\frac{d^2\Theta}{d\xi^2}\frac{1}{d\Theta(\tau(\chi))/d\xi}\frac{1}{K(\chi)}\frac{d\Theta(\tau(\chi))}{d\xi}=\frac{1}{K(\chi)}\frac{d^2\Theta}{d\xi^2}.$$

Using these relationships to express the derivatives of Θ and substituting the result into (3.2.2.1), we arrive at (3.2.2.2). The proof of the theorem is complete.

The left-hand side in (3.2.2.3) contains an expression that has the physical meaning of the flux of the transferred quantity. The transformation (3.2.2.3) is known as the localization transformation and has been introduced in [3.3].

Remark 3.2.2.2 If the function $R(\chi)/K(\chi)$ has a singularity at $\chi=0$, then instead of Eq. (3.2.2.1) we should consider the equation

$$K(\chi)\left[b\frac{d\chi}{d\tau}-\frac{d}{d\tau}\left(K(\chi)\frac{d\chi}{d\tau}\right)\right]+R(\chi)=0.$$

Examples involving the construction of some exact solutions to equations of the (3.2.2.2) type and the respective partial differential equations are discussed in [3.3] (see also Section 3.8).

Let us now turn to Eq. (3.2.2.1) and assume that the source-sink function F in this equation is the ratio $R(\chi)/K(\chi)$, with R satisfying conditions (3.2.1.9). Then the solution to Eq. (3.2.2.1) can be expressed in terms of the solution to the equation of simple waves for the KPP equation by the method discussed in Theorem 3.2.2.1. Let us illustrate this pictorially. Let $\Theta(\xi)$ be the solution to the simple-wave equation

$$-b\frac{d\Theta}{d\xi}+\frac{d^2\Theta}{d\xi^2}+R(\Theta)=0. \qquad (3.2.2.4)$$

Curve *1* in Figure 3.13 represents the function $\Theta(\xi)$, the solution to Eq. (3.2.2.4). In the same figure curve *2* represents the derivative $d\Theta/d\xi$. The graph of the function $f(\Theta) = (d\Theta/d\xi)(\Theta^{-1}(\Theta))$, with

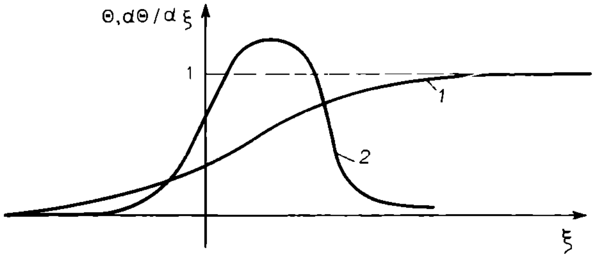


Fig. 3.13

$\Theta^{-1}(\Theta)$ the inverse of $\Theta = \Theta(\xi)$, is shown in Figure 3.14. Both Figure 3.14 and Theorem 3.2.2.1 imply that Eq. (3.2.2.1) with $F(\chi)$ equal to $R(\chi)/K(\chi)$ has a solution $\chi(\tau)$ with a semi-bounded support (this solution is shown in Figure 3.15).

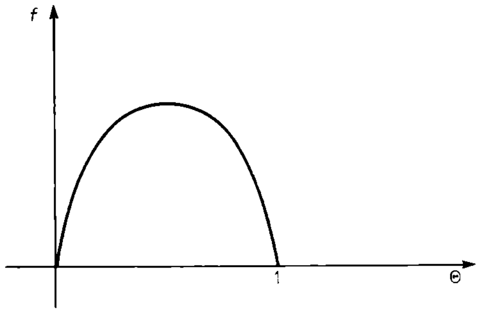


Fig. 3.14

Let us study the behavior of the function χ as $\chi \to 0$ and $\chi \to 1$. Without loss of generality, we can consider the case where the front of the weak discontinuity lies at point $\tau = 0$. As $\tau \to 0$, we have

$$\chi(\tau) = \tau^{\alpha}(\gamma_1 + o(1)), \qquad (3.2.2.5)$$

where $\gamma_1 = \gamma_1(\alpha)$ is a certain constant. Indeed, in view of (3.2.2.3), it is sufficient to prove that the function depicted in Figure 3.14 satisfies the condition $df/d\Theta|_{\Theta=0} \neq 0$. To establish the validity of

(3.2.2.5), we calculate the derivative

$$\frac{df(\Theta)}{d\Theta}\bigg|_{\Theta=0}=\frac{d}{d\Theta}\left(\frac{d\Theta}{d\xi}(\Theta^{-1}(\Theta))\right)\bigg|_{\Theta=0}$$
$$=\frac{d^2\Theta/d\xi^2}{d\Theta/d\xi}\bigg|_{\xi=-\infty}\stackrel{\text{def}}{=}l.$$

The estimate (3.2.2.5) then follows from (3.2.2.3) and the estimate (3.2.1.6) as $\xi\to-\infty$.

Let us consider the behavior of χ as $\chi\to1$. For the function $\Theta(\xi)$ the asymptotic behavior as $\xi\to+\infty$ is specified in (3.2.1.6).

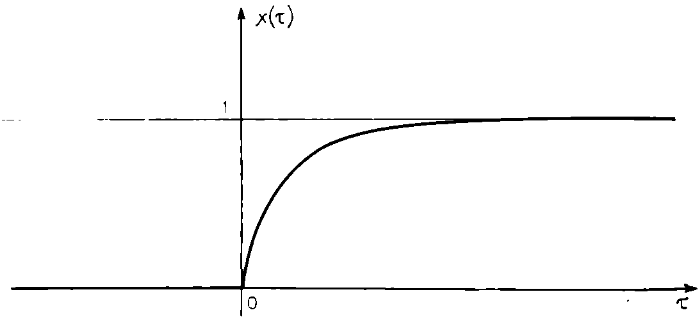


Fig. 3.15

This together with (3.2.2.3) implies that χ has the following asymptotic behavior as $\tau\to\infty$:

$$\chi=1-\exp\left(-\frac{|l_0|\tau}{K(1)}\right)+o\left(\exp\left(-\frac{|l_0|\tau}{K(1)}\right)\right).$$

The estimates for χ when $R(\Theta)$ satisfies conditions (3.2.1.5) or (3.2.1.20) can be established in a similar manner.

Summing up the above statements we arrive at an important theorem concerning the solutions to Eq. (3.2.2.1):

Theorem 3.2.2.2 *Let the functions $F(\chi)$ and $K(\chi)$ in Eq.* (3.2.2.1) *have the form $F(\chi)=\chi^qG(\chi)$, $q>0$, $k>1$, and $K(\chi)=k\rho(\chi)\chi^{k-1}$, with the functions $\rho(\chi)$ and $G(\chi)$ being continuously differentiable for $\chi\geqslant0$, $G(\chi)\neq0$ for $\chi\in[0,1)$, $G(1)=0$, and $dF/d\chi\mid_{\chi=1}\neq0$. Then there exists a monotone continuous nonnegative solution to Eq.* (3.2.2.1) *that has a semi-bounded support ($\chi(\tau)\equiv0$ for $\tau<0$), satisfies the condition $d\chi^k/d\tau\mid_{\chi=0}=0$, and is such that*

(a) *if* $k+q>2$, $q\geqslant 1$, *and* $G(\chi)>0$ *for* $\chi\in[0,1)$, *then*

$$\chi(\tau)=O(\tau^{1/(k-1)}\ \text{as}\ \tau\to 0,$$
$$\chi(\tau)=1-\exp\left(-\frac{l_0\tau}{\rho(1)k}\right)+o\left(\exp\left(-\frac{l_0\tau}{\rho(1)k}\right)\right)\ \text{as}\ \tau\to\infty,$$
$$l_0=-b/2+\sqrt{b^2/4-dR/d\Theta\,|_{\Theta=1}},$$

where $b=b_0$ *is the Zeldovich constant in Zeldovich equation*

$$b_0\frac{d\Theta}{d\xi}-\frac{d^2\Theta}{d\xi^2}-R(\Theta)=0;$$

(b) *if* $k+q=2$, $q<1$ *and* $G(\chi)>0$ *for* $\chi\in[0,1)$, *then*

$$\chi(\tau)=O(\tau^{1/(k-1)})=O(\tau^{1/(1-q)})\ \text{as}\ \tau\to 0,$$
$$\chi(\tau)=1-\exp(-l_3\tau)+o(\exp(-l_3\tau))\ \text{as}\ \tau\to\infty,$$

where

$$l_3=(-b/2+\sqrt{b^2/4+|\,dR/d\Theta\,|_{\Theta=1}|})/(k\rho(1)),$$
$$b>2\sqrt{dR/d\Theta\,|_{\Theta=0}};$$

(c) *if* $k+q>2$, $q<1$, *and* $G(\chi)<0$ *for* $\chi\in[0,1)$, *then*

$$\chi(\tau)=O(\tau^{1/(1-q)})\ \text{as}\ \tau\to 0,$$
$$\chi(\tau)=1-\exp(-l_4\tau)+o(\exp(-l_4\tau))\ \text{as}\ \tau\to\infty,$$
$$l_4=(b/2-\sqrt{b^2/4-dR/d\Theta\,|_{\Theta=1}})/(k\rho(1))<0,$$
$$b\leqslant -2\sqrt{dR/d\Theta\,|_{\Theta=1}},$$

where $R(\Theta)=k\rho(\Theta)\,\Theta^{k+q-1}G(\Theta)$, $\rho(0)>0$.

Proof. The proof of the theorem follows from Eq. (3.2.2.3) and the properties of the solution to Eq. (3.2.1.3) (see also Eq. (3.2.2.2)) discussed in Section 3.2.1.

Let us consider the ordinary differential equation

$$\hat{b}N(\chi)\frac{d\chi}{d\tau}-\frac{d}{d\tau}\left(K(\chi)\frac{d\chi}{d\tau}\right)-F(\chi)=0, \tag{3.2.2.6}$$

where $K(\chi)$ and $N(\chi)$ are positive functions, $F(\chi)>0$ for $\chi\in(0,1)$, $K(0)=0$, $K(1)>0$, and $K(\chi)=k\rho(\chi)\chi^{k-1}$. By $\Theta(\xi)$ we denote the smooth monotone solution to the equation

$$\hat{b}N(\Theta)\frac{d\Theta}{d\xi}-\frac{d^2\Theta}{d\xi^2}-R(\Theta)=0, \tag{3.2.2.7}$$

where $dR(\Theta)/d\Theta$ is continuous in $\Theta\in[0,1]$. There is a theorem that is similar to Theorem 3.2.2.1:

Theorem 3.2.2.3 *The transformation*

$$K(\chi)\frac{d\chi}{d\tau}=\frac{d\Theta}{d\xi}(\tau(\chi)) \tag{3.2.2.8}$$

reduces Eq. (3.2.2.7) *to Eq.* (3.2.2.6) *with* $F(\chi)=R(\chi)/K(\chi)$.

Proof. The proof is similar to that for Theorem 3.2.2.1.

The properties of the solutions to Eq. ((3.2.2.7) are specified by the following

Lemma 3.2.2.1 *Let the function* $R(\Theta)$ *in Eq.* (3.2.2.7) *have the properties specified by one of the conditions* (3.2.1.4), (3.2.1.5), *or* (3.2.1.20). *Then there exists a monotone nonnegative solution to Eq.* (3.2.2.7) *with* $\Theta\in[0,1]$ *such that*

(a) *if* $R(\Theta)$ *satisfies* (3.2.1.5), *then*

$$\Theta(\xi)\approx\exp(l\xi),\ l=\hat{b}N(0),\ \text{as}\ \xi\to-\infty,$$
$$\Theta(\xi)\approx 1-\exp(-l_0\xi)\ \text{as}\ \xi\to\infty,$$
$$l_0=-\hat{b}N(1)/2+\sqrt{\hat{b}^2N^2(1)/4+|dR/d\Theta|_{\Theta=1}|};$$

(b) *if* $R(\Theta)$ *satisfies* (3.2.1.4), *then*

$$\Theta\approx\exp(l\xi)\ \text{as}\ \xi\to-\infty,$$
$$\Theta\approx 1-\exp(-\hat{l}_3\xi)\ \text{as}\ \xi\to+\infty,$$
$$l=\hat{b}N(0)/2-\sqrt{\hat{b}^2N^2(0)/4-dR/d\Theta|_{\Theta=0}},$$
$$\hat{l}_3=-\hat{b}N(1)/2+\sqrt{\hat{b}^2N^2(1)/4+|dR/d\Theta|_{\Theta=1}|},$$

where $\hat{b}>2N^{-1}(0)\sqrt{dR/d\Theta|_{\Theta=0}}$;

(c) *if* $R(\Theta)$ *satisfies* (3.2.1.20) *and has an algebraic branch point at* $\Theta=0$, *or* $R(\Theta)\sim\Theta^{\beta}$, $\beta>1$, *then*

$$\Theta(\xi)\approx O(|\xi|^{-1/(\beta-1)})\ \text{as}\ \xi\to-\infty,$$
$$\Theta(\xi)\approx 1-\exp(l_4\xi)\ \text{as}\ \xi\to+\infty,$$
$$l_4=\hat{b}N(1)/2-\sqrt{\hat{b}^2N^2(1)/4-dR/d\Theta|_{\Theta=1}}<0,$$

where $\hat{b}<-2N^{-1}(1)\sqrt{dR/d\Theta|_{\Theta=1}}$.

Proof. To prove this lemma, one can employ the line of reasoning used in Section 3.2.1.

Summing up the aforesaid, we arrive at the following theorem on the properties of the solution to Eq. (3.2.2.6).

Theorem 3.2.2.4 *Let the functions* $F(\chi)$ *and* $K(\chi)$ *in Eq.* (3.2.2.6) *have the form* $F(\chi)=\chi^qG(\chi)$, $q>0$, $k>1$, $dF/d\chi|_{\chi=1}\neq 0$, *and* $K(\chi)=k\rho(\chi)\chi^{k-1}$, *and let the functions* $\rho(\chi)$, $G(\chi)$, *and* $N(\chi)$ *be continuously differentiable for* $\chi\geqslant 0$, $G(\chi)\neq 0$ *for*

$\chi \in [0, 1)$, *and* $G(1) = 0$. *Then there exists a continuous monotone nonnegative solution to Eq.* (3.2.2.6) *that has a semi-bounded support* ($\chi(\tau) \equiv 0$ *for* $\tau < 0$), *satisfies the conditions* $d\chi^k/d\tau \mid_{\chi=0} = 0$, *and is such that*

(a) *if* $k + q > 2$, $q > 1$, *and* $G(\chi) > 0$ *for* $\chi \in [0, 1]$, *then*

$$\chi(\tau) \approx O(\tau^{1/(k-1)}) \quad \textit{as } \tau \to 0,$$

$$\chi(\tau) \approx 1 - \exp(-l_0\tau/k\rho(1)) \quad \textit{as } \tau \to +\infty;$$

(b) *if* $k+q=2$, $q<1$, *and* $G(\chi)>0$ *for* $\chi \in [0, 1)$, *then*

$$\chi(\tau) \approx O(\tau^{1/(k-1)}) = O(\tau^{1/(1-q)}) \quad \textit{as } \tau \to 0,$$

$$\chi(\tau) \approx 1 - \exp(-\tilde{l}_3\tau), \quad \tilde{l}_3 = \hat{l}_3/(k\rho(1)), \quad \textit{as } \tau \to +\infty$$

(*here* $\hat{b} > 2N^{-1}(0)\sqrt{dR/d\Theta\,|_{\Theta=1}}$);

(c) *if* $k+q>2$, $q<1$, *and* $G(\chi)<0$ *for* $\chi \in [0, 1)$, *then*

$$\chi(\tau) \approx O(\tau^{1/(1-q)}) \quad \textit{as } \tau \to 0,$$

$$\chi(\tau) \approx 1 - \exp(-\hat{l}_4\tau) \quad \textit{as } \tau \to +\infty,$$

$$\hat{l}_4 = l_4/(k\rho(1)), \quad b \leqslant -2N^{-1}(1)\sqrt{dR/d\Theta\,|_{\Theta=1}},$$

where $R(\Theta) = k\rho(\Theta)\,\Theta^{k+q-1}G(\Theta)$, $\rho(0) > 0$.

Proof. The proof of this theorem is similar to that of Theorem 3.2.2.2.

3.3 A Time-dependent Model of Thermal Oxidation of Silicon

In this section we consider a time-dependent model of thermal oxidation of silicon for the case where the effective Debye shielding length depends on the coordinate in the region occupied by a growing oxide layer.[3]

A model of thermal oxidation of silicon has been described in Section 3.1. It consisted of the equation for diffusion of oxygen in a solid and of a Poisson equation describing the potential generated by the space charge formed in the Si-SiO$_2$ interface. The mathematical statement of the problem is

$$\varepsilon^3 \frac{\partial v}{\partial t} - \varepsilon^2 \frac{\partial^2 v}{\partial y^2} + \varepsilon^2 \frac{\partial}{\partial y}\left(v \frac{\partial \varphi}{\partial y}\right) = 0, \tag{3.3.0.1}$$

$$\varepsilon^3 \frac{\partial^2 \varphi}{\partial y^2} = v(y, t, \varepsilon) \sinh \varphi.$$

[3] See also [3.3-3.6].

The movement of the $Si\text{-}SiO_2$ interface is specified by the equation

$$\frac{dy_0}{dt} = C^* \left(-\frac{\partial y}{\partial y} + v \frac{\partial \varphi}{\partial y} \right) \Big|_{y=y_0(t)}, \tag{3.3.0.2}$$

where C^* is a given constant. The boundary conditions have the form

$$\begin{aligned} &v(0, t, \varepsilon) = 1, \quad v(y_0(t), t, \varepsilon) = 0, \\ &\varphi(0, t, \varepsilon) = 0, \quad \varphi(y_0(t), t, \varepsilon) = \Phi_0 < 0. \end{aligned} \tag{3.3.0.3}$$

Equations (3.3.0.1) and (3.3.0.2) together with the boundary conditions (3.3.0.3) constitute a complete mathematical model of the process of thermal oxidation of silicon in the case where the kinetics of the thermal oxidation is close to steady-state.

We will seek the asymptotic solution of the problem formulated above in the form

$$\begin{aligned} v(y, t, \varepsilon) = &[\varepsilon W_0(y, t, \tau) + \varepsilon^2 W_1(y, t, \tau) \\ &+ O(\varepsilon^3)]_{\tau=S(y,t,\varepsilon)/\varepsilon}, \end{aligned} \tag{3.3.0.4}$$
$$\varphi(y, t, \varepsilon) = [F_0(\tau) + \varepsilon F_1(y, t, \tau) + O(\varepsilon^2)]_{\tau=S(y,t,\varepsilon)/\varepsilon},$$

where $S(y, t, \varepsilon) = S_0(y, t) + \varepsilon S_1(y, t, \varepsilon)$, with $S_0(y, t) = \beta(t)(y_0(t) - y)$, $\beta(t)$, and $S_1(y, t, \varepsilon)$ being smooth functions.

Here is a theorem that describes the structure of the asymptotic solution and provides a method for solution.

Theorem 3.3.0.1 *Problem* (3.3.0.1)-(3.3.0.3) *has an asymptotic solution of the form* (3.3.0.4). *The function $W_0(\tau)$ is then determined thus:*

$$W_0(\tau) = (y_0(t))^{2/3}[\exp(F_0(\tau))] \int_0^\tau \exp(-F_0(\tau')d\tau', \tag{3.3.0.5}$$

with $W_0(\tau) \sim \tilde{C}\tau$ as $\tau \to +\infty$, and the function $F_0(\tau)$ is the solution to the boundary value problem

$$\frac{d^2F_0}{d\tau^2} = (\exp F_0)(\sinh F_0)\left(\int_0^\tau \exp(-F_0(\tau'))\, d\tau' \right),$$

$$F_0|_{\tau=0} = \Phi_0, \quad F_0|_{\tau\to\infty} = 0.$$

The function S has the form $S(x, t) = (y(t) - y) + \varepsilon S_1(x, t)$. The movement of the boundary of the region (the interface) is specified by a parabolic law, that is,

$$y_0(t) = \sqrt{2C^*t} + y_0(0). \tag{3.3.0.6}$$

Proof. We substitute solution (3.3.0.4) in (Eq. (3.3.0.1) and nullify the coefficient of the same powers of ε. We arrive at a system of

equations of the type

$$\varepsilon^1:\quad -\frac{\partial^2 W_0}{\partial\tau^2}\left(\frac{\partial S_0}{\partial y}\right)^2+\frac{\partial}{\partial\tau}\left(W_0\frac{\partial F_0}{\partial\tau}\right)\left(\frac{\partial S_0}{\partial y}\right)^2=0,$$

$$\frac{\partial^2 F_0}{\partial\tau^2}\left(\frac{\partial S_0}{\partial y}\right)^2=W_0\sinh F_0; \tag{3.3.0.7}$$

$$\begin{aligned}\varepsilon^2:\quad & -2\frac{\partial^2 W_0}{\partial\tau^2}\frac{\partial S_0}{\partial y}\frac{\partial S_1}{\partial y}-\frac{\partial^2 W_1}{\partial\tau^2}\left(\frac{\partial S_0}{\partial y}\right)^2\\ &+2\frac{\partial}{\partial\tau}\left(W_0\frac{\partial F_0}{\partial\tau}\right)\frac{\partial S_0}{\partial y}\frac{\partial S_1}{\partial y}\\ &+\left[\frac{\partial}{\partial\tau}\left(W_0\frac{\partial F_1}{\partial\tau}\right)+\frac{\partial}{\partial\tau}\left(W_1\frac{\partial F_0}{\partial\tau}\right)\right]\left(\frac{\partial S_0}{\partial y}\right)^2\\ &=\frac{\partial^2 W_0}{\partial y\,\partial\tau}\frac{\partial S_0}{\partial y}+\frac{\partial}{\partial y}\left(W_0\frac{\partial F_0}{\partial\tau}\right)\frac{\partial S_0}{\partial y},\end{aligned}$$

$$\begin{aligned}2\frac{\partial^2 F_0}{\partial\tau^2}\frac{\partial S_0}{\partial y}\frac{\partial S_1}{\partial y}+\frac{\partial^2 F_1}{\partial\tau^2}\left(\frac{\partial S_0}{\partial y}\right)^2&\\ =W_0F_1\cosh F_0+W_1\sinh F_0.&\end{aligned} \tag{3.3.0.8}$$

The boundary conditions for system (3.3.0.7) are

$$\begin{aligned}&W_0\,|_{\tau=0}=0, && W_0\,|_{\tau\to\infty}=1;\\ &F_0\,|_{\tau=0}=\Phi_0<0, && F_0\,|_{\tau\to\infty}=0.\end{aligned} \tag{3.3.0.9}$$

Let us consider system (3.3.0.7). In view of the first condition in (3.3.0.9) and the fact that $S\mid_{y=y_0(t)}=0$ we get

$$W_0(\tau)=\tilde{C}\exp(F_0(\tau))\int_0^\tau\exp(-F(\tau'))\,d\tau', \tag{3.3.0.10}$$

where $\tilde{C}=\tilde{C}(y,t)$ is an unknown function. Susbtituting this expression for function (3.3.0.10) into the second equation in (3.3.0.7), we arrive at an equation for finding F_0:

$$\frac{\beta^2}{\tilde{C}}\frac{\partial^2 F_0}{\partial\tau^2}=(\sinh F_0)\exp(F_0(\tau))\int_0^\tau\exp(-F_0(\tau'))\,d\tau',$$

$$F_0\,|_{\tau=0}=\Phi_0<0,\ F_0\,|_{\tau\to\infty}=0. \tag{3.3.0.11}$$

Let us study the behavior of (3.3.0.10) as $\tau\to\infty$. Obviously,

$$W_0\sim\tilde{C}\tau=\tilde{C}\beta(t)(y_0(t)-y)+O(\varepsilon). \tag{3.3.0.12}$$

Note that at $y=y_0(t)$ the function $S(y,t,\varepsilon)/\varepsilon$ grows without limit as $\varepsilon\to 0$. For this reason the boundary conditions of the initial problem correspond at $y=y_0(t)$ to the boundary conditions as

$\tau \to \infty$ belonging to the system of equations (3.3.0.7), (3.3.0.8). Thus, (3.3.0.9) yields

$$\tilde{C}\beta y_0 \ (t) = 1.$$

Hence, $\tilde{C} = \tilde{C}\,(t)$. Without loss of generality we can put $\beta^2 = \tilde{C}$, since in the problem studied here, in view of (3.3.0.11), a change in

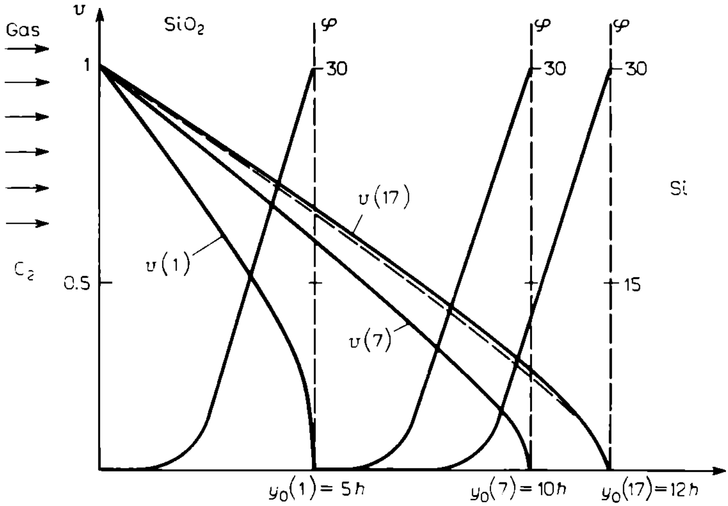


Fig. 3.16

β leads to the renormalization of S and has no effect on the final result (see [3.3]). The final result is

$$\tilde{C} = (y_0(t))^{-2/3}, \quad \beta(t) = (y_0(t))^{-1/3}.$$

Now, from the flux balance equation (3.3.0.2) we can find $y_0\,(t)$, which determines the movement of the Si-SiO_2 interface. Indeed, in view of (3.3.0.2), we have, as $\tau \to \infty$,

$$\frac{dy_0}{dt} \quad C^* \left(\tilde{C}\beta + \beta W_0 \frac{\partial F_0}{\partial \tau}\right),$$

but the last term decreases exponentially as $\tau \to \infty$, whence

$$dy_0/dt \quad C^*/y_0\,(t), \quad y_0\,(t) = \sqrt{2C^*t} + y_0\,(0).$$

As a result of a number of manipulations, system (3.3.0.8) can be reduced to the following system of equations for functions W_1

and F_1:

$$-\frac{\partial^2 W_1}{\partial \tau^2}+\frac{\partial}{\partial \tau}\left(W_0\frac{\partial F_1}{\partial \tau}+W_1\frac{\partial F_0}{\partial \tau}\right)$$
$$=2\frac{\partial^2 W_0}{\partial \tau^2}\frac{\partial S_1}{\partial y}-2\frac{\partial}{\partial \tau}\left(W_0\frac{\partial F_0}{\partial \tau}\right)\frac{\partial S_1}{\partial y}, \quad (3.3.0.13)$$
$$\frac{\partial^2 F_1}{\partial \tau^2}\beta^2-W_0F_1\cosh F_0-W_1\sinh F_0=-2\frac{\partial^2 F_0}{\partial \tau^2}\beta\frac{\partial S_1}{\partial y}.$$

The boundary conditions are

$$\begin{aligned} W_1|_{\tau=0}&=0, \quad W_1|_{\tau\to\infty}=0,\\ F_1|_{\tau=0}&=0, \quad F_1|_{\tau\to\infty}=0. \end{aligned} \quad (3.3.0.14)$$

Refining the solution, that is, the solution to higher-order equations, does not lead to qualitatively new results, and we will not undertake it here.

In conclusion we note that the above result (the law of motion of the Si-SiO_2 interface) is in good agreement with the entire body of experimental data and with the existing empirical models. In Figure 3.16 we have collected the results of numerical calculations of problem (3.3.0.1)-(3.3.0.3) (see also Figure 3.1), where the dotted curves correspond to the principal terms W_0 and F_0 in the asymptotic solution specified in Theorem 3.3.0.1. All constants are listed in Table 3.3.1.

Table 3.3.1

φ_0	Δt	h	C^*	Pe	Figure
−20	0.02	0.025	1	10^{-4}	3.1
−30	0.02	0.025	1	10^{-4}	3.16

3.4 Oxidation of Silicon in a Halogen-containing Medium

In this section we discuss a simple phenomenological model of the diffusion of chlorine in silicon. When gaseous chlorine is brought into contact with a silicon crystal, classical diffusion of chlorine into silicon occurs (on the whole, chlorine is poorly dissolved in silicon). The situation changes drastically when oxygen is admitted and the solid-phase chemical reaction starts. A pronounced maximum in the chlorine concentration is formed as a result of this process and it is localized at the Si-SiO_2 interface and moves in space. This section is devoted to the study of this process.

3.4.1 The Model Problem

In this section we define a localized asymptotic solution and build such a solution for a nonlinear equation with constant coefficients.

In the phenomenological model of silicon oxidation in a halogen-containing medium, the diffusion coefficient depends on the concentration gradient (see Section 3.1.1.2). Such a model has been studied in [3.31].

Asymptotic solutions with a compact support for quasilinear parabolic equations have been built in [3.3]. In the one-dimensional case the left and right fronts of the weak discontinuity move in opposition to each other at the same speed, so that the solution support

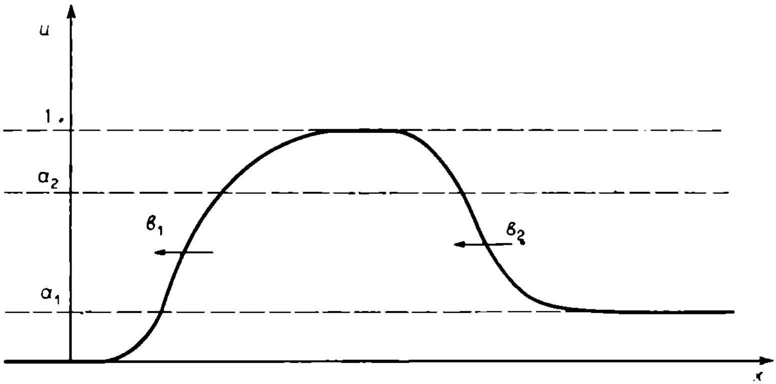


Fig. 3.17

spreads or contracts with respect to a fixed point (the "center"). Multidimensional localized asymptotic solutions with a compact support have a similar property. In this section we build one-sided localized asymptotic solutions for a quasilinear parabolic equation; the maxima of these solutions propagate in space (see Figure 3.17). Such solutions are similar to solitary waves (solitons) in the theory of wave processes, but have not been previously mentioned in the literature.

Biologists [3.25] and specialists in the field of technology [3.4] have long known of the effect of self-motion of structures, that is, motion not associated with that of the external active medium. The solutions set up in this section describe the self-motion of a dissipative structure and may be used for mathematical modeling of processes observed in experiments. We will call such solutions *solitary synergets*.

Let us consider the quasilinear parabolic equation

$$Lu = \varepsilon \frac{\partial u}{\partial t} - \varepsilon^2 \frac{\partial}{\partial x}\left(\frac{\lambda(x, t)}{1+\beta \mid \partial u/\partial x \mid} \frac{\partial u}{\partial x}\right) - \gamma^2(x, t) F(u) = 0, \tag{3.4.1.1}$$

$$x \in R^1, \quad t \in [0, T],$$

in which the function $F(u)$ has four roots, $u_1 = 0$, $u_2 = a_1 > 0$, $u_3 = a_2 > 0$, $u_4 = 1$, $a_1 < a_2 < 1$, and satisfies the following conditions:

$$dF/du\,|_{u=0} > 0, \quad dF/du\,|_{u=a_1} < 0, \quad dF/du\,|_{u=a_2} > 0, \tag{3.4.1.2}$$

$$dF/du\,|_{u=1} < 0, \quad F(u) \sim u^q \,(q \geqslant 1) \text{ as } u \to 0.$$

Definition 3.4.1.1 An asymptotic solution to Eq. (3.4.1.1) of the *solitary synerget type* is a continuous function $u(x, t, \varepsilon)$ bounded below by a positive (and independent of ε) constant on a set with positive measure and such that

(a) $Lu = Q(x, t, \varepsilon)$, where Q is a bounded function (as $\varepsilon \to 0$) for which the following estimates (or bounds) are valid:

$$\left| \frac{\partial^{\delta+\mu} Q}{\partial t^\delta\, \partial x^\mu} \right| < C_{\delta\mu} \varepsilon^{2-\delta-\mu}, \quad C_{\delta\mu} = \text{const},$$

$$\mu, \delta \in Z_+, \quad \delta + \mu \leqslant [1 + 1/(k-1)]$$

(here $[A]$ stands for the integral part of number A);

(b) the transferred-quantity flux is continuous:

$$\rho \frac{\partial u^k}{\partial x}\Big|_{x \to 0_+} \to 0, \quad \frac{1}{1+\beta \mid \partial u/\partial x \mid} \frac{\partial u}{\partial x}\Big|_{|x| \to \infty} \to 0;$$

(c) $(\partial u/\partial t)/(\partial u/\partial x) > 0$ as $|\nabla u| \to \max$.

Let us study Eq. (3.4.1.1) with constant coefficients:

$$Lu = \varepsilon \frac{\partial u}{\partial t} - \varepsilon^2 \frac{\partial}{\partial x}\left(\frac{1}{1+\beta \mid \partial u/\partial x \mid} \frac{\partial u}{\partial x}\right) - F(u) = 0, \tag{3.4.1.3}$$

$$x \in R^1, \quad t \in [0, T], \quad u \geqslant 0,$$

where $F(u) = u^q G(u)$, $q \geqslant 1$.

The function $F(u)$ vanishes four times on the segment $u \in [0, 1]$ and satisfies conditions (3.4.1.2). The asymptotic solution of the solitary synerget type to Eq. (3.4.1.3) satisfies the following boundary conditions:

$$u\,|_{x \to -\infty} \to 0, \quad u\,|_{x \to +\infty} \to a_1, \quad \partial u/\partial x\,|_{x \to -\infty} \to 0. \tag{3.4.1.4}$$

The part of the solution of the solitary-synerget type corresponding to the nonzero boundary condition is constructed by employing the

solution of the simple-wave type [3.26-3.28] (see Figure 3.3), and the part corresponding to the boundary condition as $x \to +\infty$ in (3.4.1.4) is constructed on the basis of a simple wave (see [3] and Chapter 6).

When the coefficients are constant, the solution of the solitary-synerget type to Eq. (3.4.1.1) can be constructed by employing the invariant solutions to Eq. (3.4.1.3). These are determined by the solutions $\chi(\tau)$ of the ordinary differential equation

$$b\frac{d\chi}{d\tau}-\frac{d}{d\tau}\left(\frac{1}{1+\beta\,|d\chi/d\tau|}\frac{d\chi}{d\tau}\right)-F(\chi)=0. \qquad (3.4.1.5)$$

The following lemma holds true:

Lemma 3.4.1.1 *Let*

$$dF/d\chi\,|_{\chi=0}>0, \quad dF/d\chi\,|_{\chi=1}<0, \quad dF/d\chi\,|_{\chi=a_1}<0,$$

$$b_i<2\sqrt{dF/d\chi\,|_{\chi=a_2}}, \quad b_i>\frac{1}{|dF/d\chi\,|_{a_1}|}\left(|dF/d\chi\,|_{a_1}|^3-1\right),$$

$$\int_0^1 F(\chi)\,d\chi>0, \; i=1,2. \qquad (3.4.1.6)$$

For one thing, $b_1>2\sqrt{dF/d\chi\,|_{\chi=0}}$ if $q=1$ or $b=b_0$ are the constants of the Zeldovich type in the equation

$$b_0\frac{d\chi}{d\tau}-\frac{d}{d\tau}\left(\frac{1}{1+\beta\,|d\chi/d\tau|}\frac{d\chi}{d\tau}\right)-F(\chi)=0, \quad q>1.$$

Then there exists a monotone solution χ_1 to

$$b_1\frac{d\chi_1}{d\tau}-\frac{d}{d\tau}\left(\frac{1}{1+\beta\,|d\chi_1/d\tau|}\frac{d\chi_1}{d\tau}\right)-F(\chi_1)=0$$

satisfying the conditions $0\leqslant\chi_1\leqslant 1$,

$$\chi_1(\tau)=O(\exp(l\tau)), \quad l=b_1/2+\sqrt{b_1^2/4-dF/d\chi\,|_{\chi=0}},$$

as $\tau\to-\infty$,

$$\chi_1(\tau)=1-\exp(l_1\tau)+o(\exp(l_1\tau)),$$

$$l_1=-b_1/2-\sqrt{b_1^2/4-dF/d\chi\,|_{\chi=1}}, \quad \textit{as } \tau\to+\infty$$

and a monotone solution $\chi_2(\tau)$ to

$$b_2\frac{d\chi_2}{d\tau}-\frac{d}{d\tau}\left(\frac{1}{1-\beta\,(d\chi_2/d\tau)}\frac{d\chi_2}{d\tau}\right)-F(\chi_2)=0$$

satisfying the conditions $a_1\leqslant\chi_2(\tau)\leqslant 1$,

$$\chi_2(\tau)=1-\exp(l_2\tau)+o(\exp(l_2\tau)) \quad \textit{as } \tau\to-\infty.$$

$$\chi_2(\tau)=a_1+\exp(-l_3\tau)+o(\exp(-l_3\tau)) \quad \textit{as } \tau\to+\infty,$$

with

$$l_3 = b_2/2 + \sqrt{b_2^2/4 - |dF/d\chi|_{\chi=a_1}|},$$
$$l_2 = -b_2/2 + \sqrt{b_2^2/4 - |dF/d\chi|_{\chi=1}}.$$

Theorem 3.4.1.1 *The asymptotic solution of the local-solitary-synerget type to problem* (3.4.1.1), (3.4.1.2), (3.4.1.4) *has the form*

$$u_1(x, t, \varepsilon) = \left[\chi_1\left(\frac{x+b_1t}{\varepsilon}\right)\left(1 - E_1\left(\frac{x+b_1t}{\varepsilon} - m_1\right)\right) + E_1\left(\frac{x+b_1t}{\varepsilon} - m_1\right)\right]\left[\chi_2\left(\frac{x+b_2t}{\varepsilon}\right) \times \left(1 - E_2\left(\frac{x+b_2t}{\varepsilon} - m_2\right)\right) + E_2\left(\frac{x+b_2t}{\varepsilon} - m_2\right)\right], \tag{3.4.1.7}$$

where $m_i = \text{const}$, $m_i > 0$, *and the* $E_i(\xi)$ *are smooth nonnegative functions:*[4]

$$E_i(\xi) = \begin{cases} 0 & \text{if} \quad \xi \leqslant 0, \\ 1 & \text{if} \quad \xi \geqslant \xi_{0i} > 0, \quad i = 1, 2. \end{cases} \tag{3.4.1.7'}$$

This solution satisfies Eq. (3.4.1.1) *to within* $O(\varepsilon^N)$, *where the estimate* O *is uniform in variables* x *and* t, *and* N *is any positive number.*

Proof. The proof follows from Lemma 3.4.1.1.

In variables x and t the function χ_1 constitutes a wave moving to the left, while the function χ_2 describes the "tail" of solution (3.4.1.7), which tail in variables x and t is a wave moving also to the left (see Figure 3.3).

Proof of Lemma 3.4.1.1 We start by studying the first-order ordinary differential equation

$$\frac{dp}{d\chi} = [bp - F(\chi)](1 + \beta|p|)^2/p, \tag{3.4.1.8}$$

which is obtained from Eq. (3.4.1.5) by substituting $p(\chi)$ for $d\chi/d\tau$. In the p, χ plane, Eq. (3.4.1.8) has four singular points (Figure 3.18), namely,

$$A\ (0, 0),\ B\ (1, 0),\ C\ (a_1, 0)\ D\ (a_2, 0).$$

The point A (0, 0) is a nonstable node [3.32] at $q = 1$. Indeed, the eigenvalues of the linearized system of ordinary differential equations corresponding to (3.4.1.8) are

$$\lambda_{1,2} = b/2 \pm \sqrt{b^2/4 - dF(\chi)/d\chi|_{\chi=0}}.$$

By hypothesis,

$$b \geqslant 2\sqrt{dF/d\chi|_{\chi=0}},$$

[4] These functions are used to match χ_1 and χ_2 with unity.

which implies that the eigenvalues $\lambda_{1,2}$ are real and of the same sign. The corresponding eigenvectors have the form

$$v_1 = (\lambda_1, 1),\ v_2 = (\lambda_2, 1).$$

The pattern of integral curves in the neighborhood of point A for $q \geqslant 1$ is depicted in Figure 3.18. On curve *1* we have $p = F(\chi)$ and

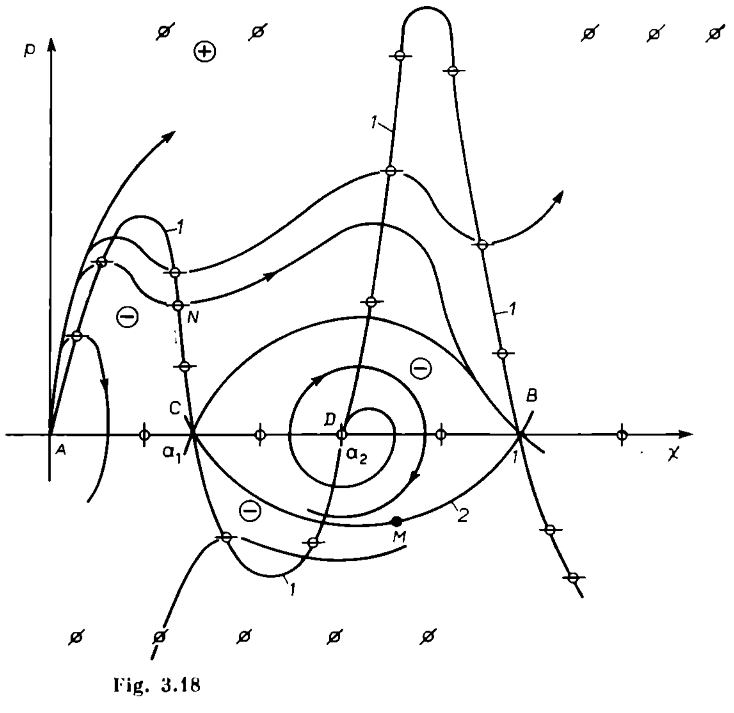


Fig. 3.18

$dp/d\chi = 0$. For $q > 1$ point A is a saddle-node [3.33], and the directions at which the integral curves emerge from this point coincide.

Point $C(a_1, 0)$ is a saddle, since the eigenvalues have different signs:

$$\lambda_{1,2} = b/2 \pm \sqrt{b^2/4 + |dF/d\chi|_{\chi=a_1}|}\ .$$

The corresponding eigenvectors are

$$v_1 = (\lambda_1, 1),\ v_2 = (\lambda_2, 1),\ \lambda_2 < 0.$$

The integral curve *2* must pass in the neighborhood of point C as shown in Figure 3.18 and majorize curve *1*, on which $p =$

$[dF/d\chi \mid_{\chi=a_1}] (a_1 - \chi) + o(a_1 - \chi)$. Thus,

$$\lambda_2^{-1} \leqslant dF/d\chi \mid_{\chi=a_1},$$

from which it follows that

$$b > |dF/d\chi \mid_{\chi=a_1}|^{-1} (|dF/d\chi|_{\chi=a_1}|^3 - 1).$$

At point $D(a_2, 0)$, which is an unstable focus [3.33], the eigenvalues and eigenvectors are

$$\lambda_{1,2} = b/2 \pm \sqrt{b^2/4 - dF/d\chi \mid_{\chi=a_2}},$$

$$v_1 = (\lambda_1, 1), \quad v_2 = (\lambda_2, 1).$$

By the hypothesis of Lemma 3.4.1.1,

$$b < 2\sqrt{dF/d\chi \mid_{\chi=a_2}}$$

and, hence, the eigenvalues are complex-valued, with $\operatorname{Re} \lambda_{1,2}$ positive.

The point $B(1, 0)$ is a saddle point.

To study the behavior of the integral curves at infinity, we extend the Euclidean plane Π to the RP^2 projective plane referred to homogeneous coordinates X, Y, Z. We will restrict our discussion to the case $F \sim \nu_0 \chi^q$, $q \geqslant 4$, as $\chi \to \infty$, with $\nu_0 = \text{const} < 0$.

Equation (3.4.1.8) in the X, Y, Z coordinates has the form [3.33]

$$\begin{vmatrix} dX & dY & dZ \\ X & Y & Z \\ P^* & Q^* & 0 \end{vmatrix} = 0, \tag{3.4.1.9}$$

where $P^* = YZ^{q-1}$ and $Q^* = YZ^{q-1} - \nu_0 X^q$.

Equation (3.4.1.9) has the form

$$-ZQ^* dX + ZP^* dY + (XQ^* - YP^*)\, dZ = 0. \tag{3.4.1.10}$$

If we nullify the expressions that are the cofactors of dX, dY, and dZ, we will find the coordinates of the singular point Φ at infinity:

$$X = 0, \quad Y = 1, \quad Z = 0.$$

At the singular point Φ we must go over to the coordinates

$$x = X/Y, \quad z = Z/Y.$$

Putting $Y = 1$, $dY = 0$, $X = x$, and $Z = z$ in (3.4.1.10), we arrive at the following system of equations

$$\frac{dz}{d\mu} = z^q b - \nu_0 z x^q,$$

$$\frac{dx}{d\mu} = xbz^{q-1} - \nu_0 x^{q+1} - z^{q-1}.$$

In the neighborhood of point $z = 0$, $x = 0$ we have $dz/dx = z/x$, that is, point Φ is a saddle-node point. We can calculate the first term in the power expansion of z (which specifies the direction of entrance into the singular point):

$$z = \left(-\frac{2\nu_0}{q+1}\right)^{1/(q-1)} x^{(q+1)/(q-1)} + \dots .$$

The curve ANB in Figure 3.18 corresponds to solution χ_1 to Eq. (3.4.1.5). This completes the proof of Lemma 3.4.1.1. The solutions χ_1 and χ_2 are then used to build the solution discussed in Theorem 3.4.1.1.

Remark 3.4.1.1 A priori it is not known whether or not there exists a constant b that satisfies all the conditions of Lemma 3.4.1.1. When $q = 1$, the conditions in the neighborhood of singular points assume the form of inequalities and there exists a certain arbitrariness in the choice of constant b for each case. Numerical integration of Eq. (3.4.1.5) has corroborated the possibility of choosing a constant b for which all the conditions of the lemma are met.

For $q > 1$, the transformation (3.2.2.3) can be used to reduce Eq. (3.4.1.5) to an equation of the (3.2.2.2) type, whose properties at $\Theta \simeq 0$ are close to those of the Zeldovich equation (see Section 3.2.1). It is then natural to expect that some of the conditions for b in Lemma 3.4.1.1 will be of the equation type, with the result that there is no guarantee that the emerging equations will be compatible. If, nevertheless, for $q > 1$ there exists a solution $\chi_1(\tau)$ to Eq. (3.4.1.5) that satisfies the required conditions, we can use this solution to build a solution to Eq. (3.4.1.1) of the local-solitary-synerget type.

3.4.2 Local Solitary Synergets in a Nonhomogeneous Medium

In this section we describe a method for building asymptotic solutions of the local-solitary-synerget type to equations with variable coefficients. The existence of such solutions means that the asymptotic solutions of the local-solitary-synerget type are stable under slow perturbations of the external medium.

Let us study the following quasilinear parabolic equation with a small parameter acting as coefficients of the derivatives:

$$Lu = \varepsilon\frac{\partial u}{\partial t} - \varepsilon^2\frac{\partial}{\partial x}\left[\left(\lambda(x,t)\frac{1}{1+|\partial u/\partial x|\beta}\right)\frac{\partial u}{\partial x}\right] - \gamma(x,t)\,G(u)\,u^q = 0, \qquad (3.4.2.1)$$

where $x \in R^1$, $t \in [0, T]$, $u \in [0, 1]$, $0 < \varepsilon < 1$, $q \geqslant 1$.

The function $u^q G(u)$ satisfies conditions (3.4.1.2) formulated in Section 3.4.1, and $\lambda(x, t)$ and $\gamma(x, t)$ are smooth positive functions. We start our discussion by giving an algorithm for building the asymptotic solution $u = u(x, t, \varepsilon)$ to Eq. (3.4.2.1) that satisfies the

following conditions:

$$u|_{x\to-\infty} \to 0, \quad \partial u/\partial x|_{x\to-\infty} \to 0,$$
$$u|_{x\to+\infty} \to 1. \tag{3.4.2.2}$$

A solution that satisfies conditions (3.4.2.2) is similar to the solution

$$u_1(x, t, \varepsilon) = \chi_1\left(\frac{x+b_1 t}{2}\right)$$

to Eq. (3.4.1.1). The following theorem holds true:

Theorem 3.4.2.1 *Suppose that conditions of the hypothesis of Lemma* 3.4.1.1 *are met. Then the asymptotic solution to Eq.* (3.4.2.1) *satisfying conditions* (3.4.2.2) *exists and has the form*

$$\begin{aligned} u_1(x, t, \varepsilon) &= [\chi_1(\tau) + \varepsilon W_1(t, \tau)]|_{\tau=S_1/\varepsilon} \\ &\quad \times \left(1 - E_1\left(\frac{\beta_1(x+\varphi_1)}{\varepsilon} - m\right)\right) \\ &\quad + E_1\left(\frac{\beta_1(x+\varphi_1)}{\varepsilon} - m_1\right), \end{aligned} \tag{3.4.2.3}$$

where $\delta_1 = \text{const} > 1$ *and* $m_1 \leqslant \max\limits_{t\in[0,T]} (\beta_1^2(\varepsilon\delta_1\beta_{11}))$.

The function $S_1(x, t, \varepsilon)$ *has the form*

$$\begin{aligned} S_1(x, t, \varepsilon) &= \beta_1(t)(x + \varphi_1(t)) + \beta_{11}(t)(x + \varphi_1(t))^2 \\ &\quad + \varepsilon S_{10}(x, t), \end{aligned}$$

with the functions $\beta_1(t)$ *and* $\varphi_1(t)$ *determined by the system of equations*

$$\frac{\beta^2\lambda(-\varphi_1, t)}{\gamma(-\varphi, t)} = 1, \quad \beta_1\frac{d\varphi_1}{dt} = \gamma(-\varphi_1, t)\, b_1, \tag{3.4.2.4}$$

and the function $\beta_{11}(t)$ *defined thus*:

$$\begin{aligned} \beta_{11} = -\Big[\beta_1 \frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi} l^2 - \frac{d\beta_1}{dt}\frac{l}{\beta_1} \\ + \frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi_1} \beta_1^{-1}\Big] [2\lambda(-\varphi_1, t)\, l\,(2l - b_1)]^{-1}. \end{aligned} \tag{3.4.2.5}$$

The function $E_1(\xi)$ *is defined in Eq.* (3.4.1.7′), *and* $\chi_1(\tau)$ *is the solution to Eq.* (3.4.1.5). *The function* S_{10} *is determined by solving the equation*

$$\begin{aligned} -\frac{\partial S_{10}}{\partial t} + 2\lambda(-\varphi_1, t)\,\beta_1\frac{\partial S_{10}}{\partial x}\, l + \beta_1\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi_1} l \\ + 2\beta_{11}\lambda(-\varphi_1, t) = 0. \end{aligned}$$

Finally, the function W_1 *is defined in*

Lemma 3.4.2.1 *The function $W_1(t, \tau)$ has the form*

$$W_1 = C_1 \frac{d\chi_1}{d\tau} + \frac{1}{\beta_1^2 \lambda(-\varphi_1, t)} \frac{d\chi_1}{d\tau}$$
$$\times \int_{-\infty}^{\tau} V \left(\frac{d\chi_1}{d\tau'}\right)^{-2} \left[\int_{-\infty}^{\tau'} \frac{1}{V} f_1 \left(1+\beta \frac{d\chi_1}{d\xi}\right)^2 \frac{d\chi_1}{d\xi} d\xi\right] d\tau', \tag{3.4.2.6}$$

where

$$V = \exp\left\{b_1 \int \left(1+\beta \frac{d\chi_1}{d\tau}\right)^2 d\tau\right\},$$
$$f_1 = \left[\beta_1 (\tau - S_{10}) \frac{\partial\lambda}{\partial x}\bigg|_{x=-\varphi_1}\right.$$
$$\left. + 2\beta_1 \lambda(-\varphi_1, t) \left\{\frac{2\beta_{11}}{\beta_1}(\tau - S_{10}) + \frac{\partial S_{10}}{\partial x}\right\}\right]$$
$$\times \frac{\partial}{\partial\tau}\left(\frac{1}{1+\beta(d\chi_1/d\tau)} \frac{d\chi_1}{d\tau}\right)$$
$$- \frac{d\chi_1}{d\tau}\left[\frac{\partial S_{10}}{\partial t} + \left(2\beta_{11}\frac{d\varphi_1}{dt} + \frac{d\beta_1}{dt}\right)\frac{\tau - S_{10}}{\beta_1}\right]$$
$$+ \frac{\partial\gamma}{\partial x}\bigg|_{x=-\varphi_1} \frac{\tau - S_{10}}{\beta_1} F(\chi_1) + \frac{2\beta_{11}\lambda(-\varphi_1, t)}{1+\beta(d\chi_1/d\tau)} \frac{d\chi_1}{d\tau}.$$

The following estimates hold true:

$$W_1 \simeq O[\tau \exp(2l\tau) + \exp(l\tau)] \text{ as } \tau \to -\infty.$$
$$W_1 \simeq O(\tau^2 \exp(l_1\tau)) \text{ as } \tau \to +\infty. \tag{3.4.2.6'}$$

Let us now prove the above-stated assertions.

Proof. We construct the monotone solution $u(x, t, \varepsilon)$ that satisfies the conditions

$$u|_{x\to\infty} \to 1, \quad u|_{x\to-\infty} \to 0.$$

In view of the monotonicity of $u(x, t, \varepsilon)$ we must put $|\partial u/\partial x| = \partial u/\partial x$ in Eq. (3.4.2.1). Let us substitute $u(x, t, \varepsilon) = W_0(S/\varepsilon) + \varepsilon W_1(S/\varepsilon, t)$ into Eq. (3.4.2.1). The result is

$$Lu = \left\{\beta_1 \frac{d\varphi_1}{dt}\frac{\partial W_0}{\partial\tau} - \beta_1^2 \lambda(-\varphi_1, t)\frac{\partial}{\partial\tau}\left[\frac{1}{1+\beta(\partial W_0/\partial\tau)}\frac{\partial W_0}{\partial\tau}\right]\right.$$
$$\left. - \gamma(-\varphi_1, t) F(W_0)\right\} + \varepsilon\left\{-\frac{\partial}{\partial\tau}\left(\frac{1}{1+\beta(\partial W_0/\partial\tau)}\frac{\partial W_0}{\partial\tau}\right)\frac{\partial\lambda}{\partial x}\bigg|_{x=-\varphi}\beta_1\right.$$
$$\times (\tau - S_{10}) - 2\beta_1\left(\frac{2\beta_{11}}{\beta_1}(\tau - S_{10}) + \frac{\partial S_{10}}{\partial x}\right)\lambda(-\varphi_1, t)$$
$$\times \frac{\partial}{\partial\tau}\left(\frac{1}{1+\beta_0(\partial W_0/\partial\tau)}\frac{\partial W_0}{\partial\tau}\right) + \beta_1^2\lambda(-\varphi_1, t)$$
$$\times \left\{\frac{\beta}{1+\beta(\partial W_0/\partial\tau)^2}\frac{\partial}{\partial\tau}\left(\frac{\partial W_0}{\partial\tau}\frac{\partial W_1}{\partial\tau}\right)\right.$$

$$+\frac{\beta^2\,\partial W_1/\partial\tau}{(1+\beta\,(\partial W_0/\partial\tau))^2}\,\frac{\partial W_0}{\partial\tau}\,\frac{\partial^2 W_0}{\partial\tau^2}-\frac{1}{1+\beta_0\,(\partial W_0/\partial\tau)}\,\frac{\partial^2 W_1}{\partial\tau^2}$$

$$-\frac{\beta\,(\partial W_1/\partial\tau)}{1+\beta\,(\partial W_0/\partial\tau)}\,\frac{\partial^2 W_0}{\partial\tau^2}\Big\}+\beta_1\,\frac{d\varphi_1}{dt}\,\frac{\partial W_1}{\partial\tau}-\gamma\,(-\varphi_1,\,t)\,\frac{dF\,(W_0)}{dW_0}\,W_1$$

$$+\left(2\beta_{11}\,\frac{d\varphi_1}{dt}+\frac{d\beta_1}{dt}\right)\frac{\tau-S_{10}}{\beta_1}\,\frac{\partial W_0}{\partial\tau}-\frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi_1}\frac{\tau-S_{10}}{\beta_1}\,F\,(W_0)$$

$$+\frac{\partial W_0}{\partial\tau}\,\frac{\partial S_{10}}{\partial t}-\frac{2\beta_{11}}{1+\beta\,(\partial W_0/\partial\tau)}\,\lambda\,(-\varphi_1,\,t)\,\frac{\partial W_0}{\partial\tau}\Big\}+O\,(\varepsilon^2). \quad (3.4.2.7)$$

Here we have allowed for the following relations:

$$\frac{\partial S_1}{\partial t}=\beta_1\,\frac{d\varphi_1}{dt}+\left(\frac{d\beta_1}{dt}+2\beta_{11}\,\frac{d\varphi_1}{dt}\right)\frac{\varepsilon\,(\tau-S_{10})}{\beta_1}$$

$$+\varepsilon\,\frac{\partial S_{10}}{\partial t}+O\,((\varepsilon\tau)^2)\,|_{\tau=S/\varepsilon},$$

$$\frac{\partial S_1}{\partial x}=\beta_1+\frac{2\beta_{11}}{\beta_1}\,\varepsilon\,(\tau-S_{10})+\varepsilon\,\frac{\partial S_{10}}{\partial x}+O\,((\varepsilon\tau)^2)\,|_{\tau=S_1/\varepsilon}$$

$$(3.4.2.8)$$

$$(x+\varphi_1)=\frac{\varepsilon\,(\tau-S_{10})}{\beta_1}+O\,((\varepsilon\tau)^2)\,|_{\tau=S_1/\varepsilon},$$

$$\frac{\partial^2 S_1}{\partial x^2}=2\beta_{11}+O\,(\varepsilon\tau)\,|_{\tau=S/\varepsilon},$$

$$\lambda\,(x,\,t)=\left[\lambda\,(-\varphi_1,\,t)+\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi_1}\frac{\varepsilon\,(\tau-S_{10})}{\beta_1}\right.$$

$$\left.+O\,((\tau\varepsilon)^2)\right]\Big|_{\tau=S_1/\varepsilon},$$

$$\gamma\,(x,\,t)=\left[\gamma\,(\quad\varphi_1,\,t)+\frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi_1}\frac{\varepsilon\,(\tau-S_{10})}{\beta_1}\right.$$

$$\left.+O\,((\tau\varepsilon)^2)\right]\Big|_{\tau=S_1/\varepsilon}.$$

Since the function $W_0=\chi_1(\tau)$ is a solution to

$$b_1\,\frac{d\chi_1}{d\tau}-\frac{d}{d\tau}\left(\frac{1}{1+\beta\,(d\chi_1/d\tau)}\,\frac{d\chi_1}{d\tau}\right)-\gamma\,(-\varphi_1,\,t)\,F\,(\chi_1)=0, \quad (3.4.2.9)$$

$$\chi_1\,|_{\tau\to-\infty}\to 0,\quad \chi_1\,|_{\tau\to+\infty}\to 1,$$

and the functions $\beta_1\,(t)$ and $\varphi_1\,(t)$ satisfy system (3.4.2.4), for the relationship $Lu=O\,(\varepsilon^2)$ to be valid it is sufficient that

$$\beta_1\,\frac{d\varphi_1}{dt}\,\frac{\partial W_1}{\partial\tau}+\beta_1^2\lambda\,(-\varphi_1,\,t)\left\{\frac{\beta}{(1+\beta\,(\partial W_0/\partial\tau))^2}\,\frac{\partial}{\partial\tau}\left(\frac{\partial W_0}{\partial\tau}\,\frac{\partial W_1}{\partial\tau}\right)\right.$$

$$+\frac{\beta^2\,(\partial W_1/\partial\tau)}{(1+\beta\,(\partial W_0/\partial\tau))^2}\,\frac{\partial W_0}{\partial\tau}\,\frac{\partial^2 W_0}{\partial\tau^2}-\frac{1}{1+\beta\,(\partial W_0/\partial\tau)}\,\frac{\partial^2 W_1}{\partial\tau^2} \quad (3.4.2.10)$$

$$\left.-\frac{\beta\,(\partial W_1/\partial\tau)}{1+\beta\,(\partial W_0/\partial\tau)}\,\frac{\partial^2 W_0}{\partial\tau^2}\right\}-\gamma\,(-\varphi_1,\,t)\,\frac{dF\,(W_0)}{dW_0}\,W_1=f_1,$$

where

$$f_1 = \frac{\partial}{\partial\tau}\left(\frac{1}{1+\beta\,(\partial W_0/\partial\tau)}\,\frac{\partial W_0}{\partial\tau}\right)\beta_1(\tau - S_{10})\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi_1}$$
$$+ 2\beta_1\left(\frac{2\beta_{11}}{\beta_1}(\tau - S_{10}) + \frac{\partial S_{10}}{\partial x}\right)\lambda(-\varphi_1, t)\frac{\partial}{\partial\tau}$$
$$\times\left(\frac{\partial W_0/\partial\tau}{1+\beta\,(\partial W_0/\partial\tau)}\right) - \left(2\beta_{11}\frac{d\varphi_1}{dt} + \frac{d\beta_1}{dt}\right)\frac{\tau - S_{10}}{\beta_1}\,\frac{\partial W_0}{\partial\tau}$$
$$+ \frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi_1}\frac{\tau - S_{10}}{\beta_1}F(W_0) - \frac{\partial W_0}{\partial\tau}\,\frac{\partial S_{10}}{\partial t}$$
$$+ \frac{2\beta_{11}}{1+\beta\,(\partial W_0/\partial\tau)}\lambda(-\varphi_1, t)\frac{\partial W_0}{\partial\tau}.$$

The function $\chi_1(\xi)$ can be found by solving Eq. (3.4.2.9) and is therefore a known function of its argument ξ. The boundary conditions for Eq. (3.4.2.10) have the form

$$W_1|_{\tau\to\infty} \to 0, \quad W_1|_{\tau\to-\infty} \to 0. \tag{3.4.2.11}$$

The general solution to Eq. (3.4.2.10) has the form

$$W_1 = C_1 W_{11} + C_2 W_{12} + \frac{W_{11}}{\beta_1^2\lambda(-\varphi_1, t)}\int_{-\infty}^{\tau}\frac{V}{W_{11}^2}$$
$$\times\left(\int_{+\infty}^{\xi}\frac{\tilde f_1 W_{11}}{V}\,d\tau\right)d\xi, \tag{3.4.2.12}$$

where

$$\tilde f_1 = f_1\left[\frac{-\beta}{(1+\beta\,(d\chi_1/d\tau))^2}\,\frac{\partial\chi_1}{\partial\tau} + \frac{1}{1+\beta\,(\partial\chi_1/\partial\tau)}\right]^{-1},$$
$$W_{11} = \frac{d\chi_1}{d\tau}, \quad W_{12} = W_{11}\int_{-\infty}^{\tau}\frac{V}{W_{11}^2}\,d\tau.$$

The function V, or the Wronskian of Eq. (3.4.2.10), has the form

$$V = \exp\left\{b_1\int\left(1 + \beta\frac{d\chi_1}{d\tau}\right)^2 d\tau\right\},$$

and the function χ_1 has the following asymptotic form as $\tau \to -\infty$:

$$\chi_1 = \exp(l\tau) + \exp(2l\tau) + O(\exp(3l\tau)). \tag{3.4.2.13}$$

In view of the estimates given in Lemma 3.4.1.1, as $\tau \to -\infty$, we get

$$\tilde f_1 = O(\exp(l\tau)), \quad V \sim O(\exp(b_1\tau)), \tag{3.4.2.14}$$

and the integrals in (3.4.2.12) are divergent for arbitrary functions β_{11} and S_{10}.

Nullifying the sum of the coefficients of $\tau \exp(l\tau)$, $\tau \to -\infty$, we arrive at the equation

$$\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi_1} \beta_1 l^2 + 4\lambda(-\varphi_1, t)\beta_{11} l^2$$
$$-\left(2\beta_{11}\frac{d\varphi_1}{dt} + \frac{d\beta_1}{dt}\right)\frac{l}{\beta_1} + \left.\frac{\partial\gamma}{\partial x}\right|_{x=-\varphi_1}\frac{1}{\beta_1} = 0. \quad (3.4.2.15)$$

This leads to an equation for β_{11}:

$$\beta_{11} = -\left[\beta_1 \left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi_1} l^2 - \frac{d\beta_1}{dt}\frac{l}{\beta_1}\right.$$
$$\left. + \left.\frac{\partial\gamma}{\partial x}\right|_{x=-\varphi_1}\frac{1}{\beta_1}\right]\left[4\lambda(-\varphi_1, t)\, l^2 - \frac{2}{\beta_1}\frac{d\varphi_1}{dt}\, l\right]. \quad (3.4.2.16)$$

Let us prove that the denominator in (3.4.2.16) does not vanish. Obviously,

$$4l^2 - 2lb_1 \neq 0, \quad l = b_1/2 + \sqrt{b_1^2/4 - dF/d\chi\,|_{\chi=0}}.$$

By Lemma 3.4.1.1,

$$\sqrt{b_1^2/4 - dF/d\chi\,|_{\chi=0}} \neq 0.$$

We nullify the sum of coefficients of $\exp(l\tau)$ as $\tau \to -\infty$ and arrive at the equation

$$-l\frac{\partial S_{10}}{\partial t} + 2\lambda(-\varphi_1, t)\, l^2\beta_1\frac{\partial S_{10}}{\partial x}$$
$$-\beta_1 \left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi_1} l^2 S_{10} + \beta_1 \left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi_1} l$$
$$+ 2\beta_{11}\lambda(-\varphi_1, t) - 4\beta_{11}\lambda(-\varphi_1, t)\, S_{10} l^2$$
$$+\left[2\beta_{11}\frac{d\varphi_1}{dt} + \frac{d\beta_1}{dt}\right]\frac{lS_{10}}{\beta_1}$$
$$+\left.\frac{\partial\gamma}{\partial x}\right|_{x=-\varphi_1}\frac{S_{10}}{\beta_1} = 0. \quad (3.4.2.17)$$

Combining (3.4.2.16) and (3.4.2.17) results in an equation for S_{10}:

$$-\frac{\partial S_{10}}{\partial t} + \lambda(-\varphi_1, t)\, 2\beta_1\frac{\partial S_{10}}{\partial x} l + \beta_1 \left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi_1} l$$
$$+ 2\beta_{11}\lambda(-\varphi_1, t) = 0.$$

We continue with the study of (3.4.2.12). Estimate (3.4.2.13) implies that if Eq. (3.4.2.15) is valid, function $\tilde{f}_1$ has the following behavior as $\tau \to -\infty$:

$$\tilde{f}_1 = O(\tau \exp(2l\tau)).$$

Moreover, as $\tau \to -\infty$, we have

$$\begin{aligned} &\tilde{f}_1 \frac{d\chi_1}{d\tau} V^{-1} = O(\tau \exp\{(3l - b_1)\tau\}), \\ &V\left(\frac{d\chi_1}{d\tau}\right)^{-2} = O(\exp\{(b_1 - 2l)\tau\}). \end{aligned} \tag{3.4.2.17'}$$

The inner integral in (3.4.2.12) converges as $\tau \to -\infty$ if $3b - b_1$ is positive. In general, the following estimate holds true:

$$\int_{-\infty}^{\xi} V^{-1}\tilde{f}_1 \frac{d\chi_1}{d\tau}\, d\tau = O(\xi \exp\{(3l - b)\xi\}) \text{ as } \xi \to -\infty.$$

Let us consider the behavior of the integrand in (3.4.2.12) as $\xi \to +\infty$. By virtue of the estimate

$$\begin{aligned} &\chi_1 \sim 1 - \exp(-l_1\xi), \\ &l_1 = -b_1/2 + \sqrt{b_1^2/4 + |\, dF/d\chi_1 |_{\chi_1=1} \,|}\,, \end{aligned}$$

we have (as $\xi \to +\infty$)

$$\begin{aligned} &\tilde{f}_1 \frac{d\chi_1}{d\tau}/V = O\{\tau \exp[(-2l_1 - b_1)\tau]\}, \\ &V\left(\frac{d\chi_1}{d\tau}\right)^{-2} = O\{\exp[(b_1 + 2l_1)\tau]\}. \end{aligned} \tag{3.4.2.18}$$

By virtue of the estimates (3.4.2.18), the inner integral in (3.4.2.12) is convergent. We also have the estimate

$$\int_{\tau}^{\infty} \tilde{f}_1 V^{-1} \frac{d\chi_1}{d\tau}\, d\tau = O\{\tau \exp[(3l - b_1)\tau]\},$$

and $C_2 \equiv 0$ in (3.4.2.12). Then (3.4.2.12) assumes the form given in Lemma 3.4.2.1. The above estimates imply that

$$W_1 = \begin{cases} O(\tau^2 \exp(-l_1\tau)) & \text{as } \tau \to \infty, \\ O(\tau \exp(2l\tau)) & \text{as } \tau \to -\infty. \end{cases}$$

We have thus constructed an asymptotic solution of problem (3.4.2.1), (3.4.2.2) to within $O(\varepsilon^2)$.

At this point it is important to note that for $0 < \text{const} < S_1$ the function $\chi_1(\xi, (\tau, t, \varepsilon))\,|_{\tau=S_1/\varepsilon}$ becomes exponentially close to unity as $\varepsilon \to 0$, by virtue of Lemma 3.4.1.1. This makes it possible to set up a global asymptotic solution by matching χ_1 to unity, just as was done in Section 3.4.1 and in [3.3]. This matching is performed by employing the functions E_i. The definition (3.4.1.7′) for

the function E_1 (ξ) implies that

$$E_1\left(\frac{\beta_1(x+\varphi_1)}{\varepsilon}-m_1\right)=\begin{cases}0 \text{ if } \dfrac{\beta_1(x+\varphi_1)}{\varepsilon}\leqslant m_1,\\ 1 \text{ if } \dfrac{\beta_1(x+\varphi_1)}{\varepsilon}\geqslant\dfrac{m_1\delta_1}{\delta_2},\end{cases} \tag{3.4.2.18'}$$

wnere

$$m_1\overset{\text{def}}{=}\max_{t\in[0,\,T]}|\beta_1^2/(\varepsilon\beta_{11})|.$$

The proofs of Theorem 3.4.2.1 and Lemma 3.4.2.1 are complete.

To build an analog of the solution χ_2 given in Lemma 3.4.1.1 we must construct an asymptotic solution with the same properties as the exact solution $\chi_2\left(\frac{x+bt}{\varepsilon}\right)$ to Eq. (3.4.1.5).

Here is an algorithm that can be used to construct the asymptotic solution u (x, t, ε) to Eq. (3.4.2.1) satisfying the following conditions

$$\begin{aligned} &u(x,\,t,\,\varepsilon)\,|_{x\to-\infty}\to 1, \quad \left.\frac{\partial u}{\partial x}\right|_{x\to-\infty}\to 0,\\ &u(x,\,t,\,\varepsilon)\,|_{x\to+\infty}\to a_1, \quad \left.\frac{\partial u}{\partial x}\right|_{x\to+\infty}\to 0. \end{aligned} \tag{3.4.2.19}$$

A solution to Eq. (3.4.2.1) that satisfies conditions (3.4.2.19) is smooth. Hence, the algorithm of building an asymptotic solution to be discussed differs from the algorithm employed in Section 3.4.1. The main difference in the form of representation of the asymptotic solution lies in the choice of the representation for the "phase", the function S (x, t, ε).

The formula given in Theorem 3.4.2.1 is similar, on the one hand, to the formulas given in this and previous sections and, on the other, to the forumlas used in wave theory to build asymptotic solutions. Just as we did with (3.4.2.18′), let us introduce a continuously differentiable function E_2 that matches the solution to unity. Then we have

Theorem 3.4.2.2 *Assume that the conditions of Lemma* 3.4.1.1 *are met. Then the solution to Eq.* (3.4.2.1) *satisfying conditions* (3.4.2.19) *exists and has the form*

$$u_2(x,\,t,\,\varepsilon)=\left(W_0\left(\frac{S_2}{\varepsilon}\right)+\varepsilon W_1\left(\frac{S_2}{\varepsilon},\,t,\,\varepsilon\right)\right)E_2\left(\frac{\beta_2(x+\varphi_2)}{\varepsilon}-m_2\right)$$
$$-E_2\left(\frac{\beta_2(x+\varphi_2)}{\varepsilon}-m_2\right)+1,$$

where $m_2=\max\limits_{t\in[0,\,T]}|\beta_2^2/(\varepsilon\delta_1'\beta_{21})|$, *with* $\delta_1'=\text{const}>1$. *The function* $S_2(x,\,t,\,\varepsilon)$ *has the form*

$$S_2(x,\,t,\,\varepsilon)=\beta_2(t)(x+\varphi_2(t))+\beta_{21}(t)(x+\varphi_2(t))^2+\varepsilon S_{20}(x,\,t),$$

the functions $\beta_2(t)$ *and* $\varphi_2(t)$ *are defined through the system of equations*

$$\frac{\beta_2^2\lambda(-\varphi_2, t)}{\gamma(-\varphi_2, t)} = 1, \quad \beta_2\frac{d\varphi_2(t)}{dt} = \gamma(-\varphi_2, t)\,b_2, \qquad (3.4.2.20)$$

and the function $\beta_{21}(t)$ *is defined thus:*

$$\beta_{21} = -\left(\beta_2\frac{\partial\lambda}{\partial x}\bigg|_{x=-\varphi_2} l_2^2 - \frac{d\beta_2}{dt}\frac{l_2}{\beta_2} + \frac{\partial\gamma}{\partial x}\bigg|_{x=-\varphi_2}\beta_2^{-1}\right) \times [2\lambda(-\varphi_2, t)\,l_2(2l_2-b_2)]^{-1}. \qquad (3.4.2.21)$$

The function S_{20} *can be found by solving the equation*

$$-\frac{\partial S_{20}}{\partial t}I_3 + 2\lambda(-\varphi_2, t)\beta_2\left[\frac{\partial S_{20}}{\partial x}I_1 + \frac{2\beta_{21}}{\beta_2}(I_0 - I_1S_{20})\right] + \beta_2\frac{\partial\lambda}{\partial x}\bigg|_{x=-\varphi_2}(I_0 - I_1S_{20}) - \left(2\beta_{21}\frac{d\varphi_2}{dt} + \frac{d\beta_2}{dt}\right)\frac{1}{\beta_2}(I_2 - I_3S_{20}) + \frac{1}{\beta_2}\frac{\partial\gamma}{\partial x}\bigg|_{x=-\varphi_2}(I_5 - I_6S_{20}) + 2\beta_{21}\lambda(-\varphi_2, t)\,I_4 = 0,$$

where the integrals I_0 *to* I_6 *are calculated by the formulas* (3.4.2.30) *given below. The function* W_1 *has the form*

$$W_1 = C\frac{d\chi_2}{d\tau} - \frac{1}{\beta_2^2\lambda(-\varphi_2, t)}\frac{d\chi_2}{d\tau'}\int_{-\infty}^{\tau} V\left(\frac{d\chi_2}{d\tau'}\right)^{-2} \times \left(\int_{-\infty}^{\tau'}\frac{f_2(1-\beta(d\chi_2/d\xi))^2(d\chi_2/d\xi)}{V}\,d\xi\right)d\tau',$$

where

$$V = \exp\left\{b_1\int\left(1-\beta\frac{d\chi_2}{d\tau}\right)^2 d\tau\right\},$$

$$f_2 = \Bigg[\beta_2(\tau - S_{20})\frac{\partial\lambda}{\partial x}\bigg|_{x=-\varphi_2} + 2\beta_2\lambda(-\varphi_2, t)\left\{\frac{2\beta_{21}}{\beta_2}(\tau - S_{20}) + \frac{\partial S_{20}}{\partial x}\right\}\Bigg] \times \frac{\partial}{\partial\tau}\left(\frac{1}{1-\beta\frac{d\chi_2}{d\tau}}\frac{d\chi_2}{d\tau}\right) - \frac{d\chi_2}{d\tau}\left[\frac{\partial S_{20}}{\partial t} + \left(2\beta_{21}\frac{d\varphi_2}{dt} + \frac{d\beta_2}{dt}\right)\frac{\tau - S_{20}}{\beta_2}\right] + \frac{\partial\gamma}{\partial x}\bigg|_{x=-\varphi_2}\frac{\tau - S_{20}}{\beta_2}F(\chi_2) + \frac{2\beta_{21}\lambda(-\varphi_2, t)}{1-\beta(d\chi_2/d\tau)}\frac{d\chi_2}{d\tau}.$$

The following estimates hold true

$$W_1 = \begin{cases} O(\tau\exp(2l_2\tau) + \exp(l_2\tau)) & \textit{as } \tau \to -\infty, \\ O(\tau^2\exp(l_3\tau)) & \textit{as } \tau \to +\infty. \end{cases}$$

Proof. We will set up an analog of the monotone solution χ_2 and, therefore, we will assume that $|\partial u_2/\partial x| = -\partial u/\partial x$. We substitute the function $u(x, t, \varepsilon) = W_0(S(x, t)/\varepsilon) + \varepsilon W_1(x, t, S(x, t)/\varepsilon)$ into Eq. (3.4.2.1) and get

$$
\begin{aligned}
Lu = & \Big\{\beta_2 \frac{d\varphi_2}{dt}\frac{\partial W_0}{\partial \tau} - \beta_2^2\lambda(-\varphi_2, t)\frac{\partial}{\partial \tau}\Big[\frac{1}{1-\beta(\partial W_0/\partial\tau)}\frac{\partial W_0}{\partial \tau}\Big] \\
& - \gamma(-\varphi_2, t) F(W_0)\Big\} \\
& + \varepsilon\Big\{-\frac{\partial}{\partial\tau}\Big(\frac{1}{1-\beta(\partial W_0/\partial\tau)}\frac{\partial W_0}{\partial\tau}\Big)\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi_2}\beta_2(\tau - S_{20}) \\
& - 2\beta_2\Big(\frac{2\beta_{21}}{\beta_2}(\tau - S_{20}) + \frac{\partial S_{20}}{\partial x}\Big)\lambda(-\varphi_2, t)\frac{\partial}{\partial\tau}\Big(\frac{1}{1-\beta(\partial W_0/\partial\tau)}\frac{\partial W_0}{\partial\tau}\Big) \\
& + \beta_2^2\lambda(-\varphi_2, t)\Big[\frac{-\beta}{(1-\beta(\partial W_0/\partial\tau))^2}\frac{\partial}{\partial\tau}\Big(\frac{\partial W_0}{\partial\tau}\frac{\partial W_1}{\partial\tau}\Big) \\
& + \frac{\beta^2(\partial W_1/\partial\tau)}{(1-\beta(\partial W_0/\partial\tau))^2}\frac{\partial W_0}{\partial\tau}\frac{\partial^2 W_0}{\partial\tau^2} - \frac{1}{1-\beta(\partial W_0/\partial\tau)}\frac{\partial^2 W_1}{\partial\tau^2} \\
& + \frac{\beta(\partial W_1/\partial\tau)}{1-\beta\partial W_0/\partial\tau}\frac{\partial^2 W_0}{\partial\tau^2}\Big] + \beta_2\frac{d\varphi_2}{dt}\frac{dW_1}{d\tau} - \gamma(-\varphi_2, t)\frac{dF(W_0)}{dW_0}W_1 \\
& + \Big(2\beta_{21}\frac{d\varphi_2}{dt} + \frac{d\beta_2}{dt}\Big)\frac{\tau - S_{20}}{\beta_2}\frac{dW_0}{d\tau} - \frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi_2}\frac{\tau - S_{20}}{\beta_2}F(W_0) \\
& + \frac{\partial W_0}{\partial\tau}\frac{\partial S_{20}}{\partial t} - \frac{2\beta_{21}}{1-\beta(\partial W_0/\partial\tau)}\lambda(-\varphi_2, t)\frac{\partial W_0}{\partial\tau}\Big\} + O(\varepsilon^2). \quad (3.4.2.22)
\end{aligned}
$$

Here we have employed (3.4.2.8), with the functions β_1, β_{11}, S_{10}, and S_1 replaced by β_2, β_{21}, S_{20}, and S_2, respectively.

Since the function $W_0 = \chi_2(\tau)$ is a solution to

$$
b_2\frac{\partial\chi_2}{\partial\tau} - \frac{\partial}{\partial\tau}\Big(\frac{1}{1-\beta(\partial\chi_2/\partial\tau)}\frac{\partial\chi_2}{\partial\tau}\Big) - \gamma(-\varphi_2, t)F(\chi_2) = 0, \quad (3.4.2.23)
$$

$$
\chi_2|_{\tau\to-\infty} = 1, \quad \chi_2|_{\tau\to+\infty} = a_1 + 0,
$$

and the functions $\beta_2(t, \varepsilon)$ and $\varphi_2(t, \varepsilon)$ satisfy system (3.4.2.20) for $Lu = O(\varepsilon^2)$ to be valid it is sufficient that the following equation be satisfied:

$$
\begin{aligned}
& \beta_2\frac{d\varphi_2}{dt}\frac{\partial W_1}{\partial\tau} + \beta_2\lambda(-\varphi_2, t)\Big\{\frac{-\beta}{(1-\beta(\partial W_0/\partial\tau))^2}\frac{\partial}{\partial\tau}\Big(\frac{\partial W_0}{\partial\tau}\frac{\partial W_1}{\partial\tau}\Big) \\
& + \frac{\beta^2(\partial W_1/\partial\tau)}{(1-\beta(\partial W_0/\partial\tau))^2}\frac{\partial W_0}{\partial\tau}\frac{\partial^2 W_0}{\partial\tau^2} - \frac{1}{1-\beta(\partial W_0/\partial\tau)}\frac{\partial^2 W_1}{\partial\tau^2} \\
& + \frac{\beta}{1-\beta(\partial W_0/\partial\tau)}\frac{\partial W_1}{\partial\tau}\frac{\partial^2 W_0}{\partial\tau^2}\Big\} \\
& - \gamma(-\varphi_2, t)\frac{dF(W_0)}{dW_0}W_1 = f_2, \qquad (3.4.2.24)
\end{aligned}
$$

where

$$f_2 = \frac{\partial}{\partial \tau}\left(\frac{1}{1-\beta(\partial W_0/\partial \tau)}\frac{\partial W_0}{\partial \tau}\right)\beta_2(\tau - S_{20})\frac{\partial \lambda}{\partial x}\Big|_{x=-\varphi_2}$$
$$+2\beta_2\left(\frac{2\beta_{21}}{\beta_2}(\tau - S_{20}) + \frac{\partial S_{20}}{\partial x}\right)\lambda(-\varphi_2, t)$$
$$\times\frac{\partial}{\partial \tau}\left(\frac{1}{1-\beta(\partial W_0/\partial \tau)}\frac{\partial W_0}{\partial \tau}\right)$$
$$-\left(2\beta_{21}\frac{d\varphi_2}{dt} + \frac{\partial \beta_2}{\partial t}\right)\frac{\tau - S_{20}}{\beta_2}\frac{\partial W_0}{\partial \tau}$$
$$+\frac{\partial \gamma}{\partial x}\Big|_{x=-\varphi_2}\frac{\tau - S_{20}}{\beta_2}F(W_0) - \frac{\partial W_0}{\partial \tau}\frac{\partial S_{20}}{\partial t}$$
$$+2\beta_{21}\lambda(-\varphi_2, t)\frac{\partial W_0}{\partial \tau}\Big/\left(1-\beta\frac{\partial W_0}{\partial \tau}\right).$$

The function $\chi_2(\tau)$ can be found by solving Eq. (3.4.3.23).
The boundary conditions for Eq. (3.4.2.24) are

$$W_1|_{\tau\to\infty} \to 0, \quad W_1|_{\tau\to-\infty} \to 0. \tag{3.4.2.25}$$

The general solution to Eq. (3.4.2.24) is

$$W_1 = C_1 W_{11} + C_2 W_{12} - \frac{W_{11}}{\beta_2^2 \lambda(-\varphi_2, t)} \times \int_{-\infty}^{\tau}\frac{V}{W_{11}^2}\left(\int_{-\infty}^{\xi}\frac{\tilde f_2 W_{11}}{V}d\tau\right)d\xi, \tag{3.4.2.26}$$

where

$$\tilde f_2 = f_1\frac{1}{[1-\beta(d\chi_2/d\tau)]^2}, \quad W_{11} = \frac{d\chi_2}{d\tau}, \quad W_{12} = W_{11}\int_{-\infty}^{\tau}\frac{V}{W_{11}^2}d\tau.$$

The function V, which is Wronskian of Eq. (3.4.2.24), has the form

$$V = \exp\left\{b_2\int\left(1-\beta\frac{d\chi_2}{d\tau}\right)^2 d\tau\right\}.$$

As $\tau \to -\infty$, the function χ_2 behaves in the following manner:

$$\chi_2 \sim 1 - \exp(l_2\tau) - \exp(2l_2\tau). \tag{3.4.2.27}$$

By virtue of the estimates carried out in Lemma 3.4.1.1, as $\tau \to -\infty$, we get

$$\tilde f_2 = O(\exp(l_2\tau)), \quad V \sim O(\exp(b_2\tau)),$$

and the integrals in (3.4.2.26) with arbitrary β_{21} and S_{20} are divergent.

If we nullify the sum of coefficients of $\tau \exp(l_2\tau)$ on the right-hand side f_2 of Eq. (3.4.2.24). we arrive, as $\tau \to -\infty$, at the equation

$$\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi_2} \beta_2 l_2^2 + 4\lambda(-\varphi_2, t)\beta_{21} l_2^2$$
$$-\left(2\frac{\beta_{21}}{\beta_2}\frac{d\varphi_2}{dt} + \frac{d\beta_2}{dt}\right) l_2 + \left.\frac{\partial\gamma}{\partial x}\right|_{x=-\varphi_2} \frac{1}{\beta_2} = 0. \quad (3.4.2.28)$$

From this follows the expression (3.4.2.21) for the function β_{21}.

The denominator in (3.4.2.21) does not vanish; this can be proved in the same manner as in Theorem 3.4.2.1.

Let us continue our study of (3.4.2.26). From estimate (3.4.2.27) it follows that if (3.4.2.28) is valid, then $\tilde{f}_2$ behaves, as $\tau \to -\infty$, in the following manner:

$$\tilde{f}_2 = O(\tau \exp(2l_2\tau) + \exp(l_2\tau)).$$

Moreover, as $\tau \to -\infty$, we have

$$\tilde{f}_2 \frac{d\chi_2}{d\tau}\Big/ V = O(\tau \exp\{(3l_2 - b_2)\tau\} + \exp\{(2l_2 - b_2)\tau\}),$$
$$V\left(\frac{d\chi_2}{d\tau}\right)^{-2} = O(\exp\{(b_2 - 2l_2)\tau\}).$$

The inner integral in (3.4.2.26) converges as $\tau \to -\infty$. The following is the necessary condition for W_1 to descrease as $\tau \to +\infty$:

$$I \overset{\text{def}}{=} \int_{-\infty}^{\infty} \left\{\tilde{f}_2 \frac{d\chi_2}{d\tau}\left(1 - \beta\frac{d\chi_2}{d\tau}\right)^2\right\} V^{-1}\, d\tau = 0, \quad (3.4.2.29)$$

and $C_2 = \text{const} \equiv 0$. From (3.4.2.17) follows an equation for determining S_{20}:

$$-\frac{\partial S_{20}}{\partial t} I_3 + 2\lambda(-\varphi_2, t)\beta_2\left[\frac{\partial S_{20}}{\partial x} I_1 + 2\frac{\beta_{21}}{\beta_2}(I_0 - I_1 S_{20})\right]$$
$$+\beta_2 \left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi_2} (I_0 - I_1 S_{20})$$
$$-\left(2\beta_{21}\frac{d\varphi_2}{dt} + \frac{d\beta_2}{dt}\right)\frac{1}{\beta_2}(I_2 - I_3 S_{20})$$
$$+\frac{1}{\beta_2}\left.\frac{\partial\gamma}{\partial x}\right|_{x=-\varphi_2} (I_5 - I_6 S_{20}) + 2\beta_{21}\lambda(-\varphi_2, t) I_4 = 0,$$

where the integrals I_0 to I_6 are specified as follows:

$$I_0 = \int_{-\infty}^{\infty} N \frac{\partial}{\partial\tau}\left(\frac{1}{1-\beta(d\chi_2/d\tau)}\frac{d\chi_2}{d\tau}\right)\tau\, d\tau,$$

$$I_1 = \int_{-\infty}^{\infty} N \frac{\partial}{\partial\tau}\left(\frac{1}{1-\beta(d\chi_2/d\tau)}\frac{d\chi_2}{d\tau}\right) d\tau,$$

$$I_2 = \int_{-\infty}^{\infty} \tau N \frac{d\chi_2}{d\tau}\, d\tau,\ I_3 = \int_{-\infty}^{\infty} N \frac{d\chi_2}{d\tau}\, d\tau, \tag{3.4.2.30}$$

$$I_4 = \int_{-\infty}^{\infty} N \frac{1}{1-\beta\,(d\chi_2/d\tau)} \frac{d\chi_2}{d\tau}\, d\tau,$$

$$I_5 = \int_{-\infty}^{\infty} NF(\chi_2)\,\tau\, d\tau,\ I_6 = \int_{-\infty}^{\infty} NF(\chi_2)\, d\tau,$$

with

$$N = \left(1-\beta\frac{d\chi_2}{d\tau}\right)^2 \frac{d\chi_2}{d\tau} \exp\left\{-b_2 \int^{\tau} \left(1-\beta\frac{d\chi_2}{d\tau}\right)^2 d\tau\right\}.$$

The integral in (3.4.2.29) has the following properties:

$$I = O(\tau\exp\{(3l_2-b_2)\tau\} + \exp\{(2l_2-b_2)\tau\}) \quad \text{as } \tau\to-\infty,$$
$$I = O(\tau\exp\{(2l_3-b_2)\tau\}) \quad \text{as } \tau\to+\infty,$$

and the following estimates hold true:

$$W_1 = O(\tau\exp(2l_2\tau) + \exp(l_2\tau)) \quad \text{as } \tau\to-\infty,$$
$$W_1 = O(\tau^2\exp(l_3\tau)) \quad \text{as } \tau\to+\infty$$

(for the case where $\tau\to+\infty$ we have used estimates of χ_2; see Lemma 3.4.1.1).

Thus, we have built an asymptotic solution to problem (3.4.2.1), (3.4.2.2) to within $O(\varepsilon^2)$. We can then employ the unity decomposition technique and match the solution with unity, as in Section 3.4.1. The proof of the Theorem 3.4.2.2 is complete. The two asymptotic solutions, one built in Theorem 3.4.2.1 and the other in Theorem 3.4.2.2, can be used to construct a new asymptotic solution, the solitary synerget

$$u(x, t, \varepsilon) = u_1(x, t, \varepsilon)\, u_2(x, t, \varepsilon)$$

(see Figure 3.17).

Remark 3.4.2.1 The algorithm developed in Theorems 3.4.1.1 and 3.4.2.1 makes it possible to construct any number of terms in the asymptotics of the solution of the solitary-synerget type, that is, construct the solution to $Lu = Q_N(x, t, \varepsilon)$ (see Definition 3.4.1.1), where Q_N satisfies the following condition:

$$\left|\frac{\partial^{\delta+\mu} Q_N}{\partial t^{\delta}\partial x^{\mu}}\right| < C_{\delta_\mu}\varepsilon^{N-\delta-\mu},\ C_{\delta_\mu} = \text{const},$$

$$\delta,\ \mu\in\mathbb{Z}_+,\ \delta+\mu \leqslant \left[N+\frac{1}{k-1}\right]$$

(here $[a]$ stands for the integral part of the number a for all $N > 1$).

3.5 Models of Mass Transfer

In this section we discuss mathematical models of processes of precipitation and coprecipitation of ions. From the standpoint of the theory of differential equations these models represent boundary value problems for a system of quasilinear parabolic equations. In some cases (which are discussed in this section) the system consists of two equations: a quasilinear (or linear) parabolic equation and an ordinary differential (or algebraic) nonlinear equation. The processes described by these models are used, for example, to extract, concentrate, and separate radionuclides. Obtaining quantitative as well as qualitative estimates of the parameters of these processes has important scientific and practical significance. For one thing, a study of the laws governing the precipitation and coprecipitation of radionuclides is necessary for forecasting the distribution of the nuclides in the ecological chains, which is important in environmental protection and the health services.

Dynamic precipitation and coprecipitation of ions occur during the filtration of electrolyte solutions through a dispersion medium. The parameters may vary in time and in space in the process owing to the action of acoustic waves, a temperature gradient, and other agents.

In this section we build asymptotic solutions for the mathematical models of the physico-chemical processes of the heterogeneous and interphase distribution of ions and analyze variants of the isotherm equation for the precipitation of a metal's hydroxide both with and without aquohydrocomplex formation and the isotherm equation for precipitation of salts of polybasic acids. The results of such studies can be used in calculating precipitation chromatography [3.9] and in other fields of physical chemistry and chemical technology.

3.5.1 Time-dependent Models of Mass Transfer

Let us consider the heterogeneous process of filtration of an electrolyte solution through a dispersion medium. The pumping of the substance through the medium can be carried out at a variable rate, with the variations in the rate caused by some external agent. The electrolyte solution contains the ions of the precipitating substance (cations) and the ions of the substance-precipitator (anions), and the two enter into a chemical reaction. The ions of the precipitating substance enter into the forming precipitate. The solution is filtrated by the dispersion medium, on which a layer of the sorbent grows as a result. We will assume all along that diffusion of the precipitating substance also takes place.

Here is an equation of mass conservation in dimensionless form:

$$\frac{\partial v}{\partial t}+u(y,\ t)\frac{\partial v}{\partial y}+\frac{\varepsilon}{m}\frac{\partial \varphi}{\partial t}=\varepsilon\frac{\partial}{\partial y}\left(D(y,\ t)\ \frac{\partial v}{\partial y}\right),$$
$$\varepsilon=\mathrm{Pe}^{-1},\quad m=\mathrm{const}. \tag{3.5.1.1}$$

Here v is the dimensionless concentration of the sorbate in the solution (the mobile phase), φ the dimensionless concentration of the same in the immobile phase, $u\ (y,\ t)$ is the linear dimensionless transfer rate of the mobile phase, $D\ (y,\ t)>0$ the dimensionless quasidiffusion coefficient (which depends on the spatial coordinate and time in the case of slow perturbations of the medium), Pe the diffusion Péclet number, y the dimensionless spatial coordinate directed along the flow, and t the dimensionless temporal variable.

This equation is augmented by a simple relationship that describes the sorption kinetics in dimensionless form:[5]

$$\frac{\partial \varphi}{\partial t} = f(v, \varphi). \tag{3.5.1.2}$$

When dynamical equilibrium sets in, the kinetic equation can be replaced by the isotherm equation

$$\varphi = f(v). \tag{3.5.1.3}$$

Then the given system of equations describes equilibrium sorption. The isotherm may have, say, the form $\varphi = a - b/v^{1/\lambda}$, with a, b, and λ constants and $a - b = 1$ in all cases. In this form it is a corollary of the mass balance equations for a chemical reaction and of the law of mass action, with λ a constant equal to the ratio of the stoichiometric coefficients of the reaction. A system of equations describing precipitation processes has been derived in [3.9] and discussed in detail in [3.3].

An example of an exact solution with $u = \text{const}$ in Eq. (3.5.1.1) was discussed in Section 3.1.3. But if $u = u(y, t)$, that is, if the rate of admission or extraction of the precipitating substance varies, the self-similar solution set up in Section 3.1 does not satisfy Eqs. (3.5.1.1) and (3.5.1.3). The shape of the solution, however, is preserved (see Figures 3.5 and 3.6), which makes it possible to employ the methods developed in [3.3] to construct an approximate solution.

3.5.2 An Asymptotic Solution to the Kinetic Equation for Equilibrium Adsorption (Desorption) in the Case of Soluble Precipitates

In this section we construct an asymptotic solution to the system of parabolic equations describing the precipitation process in the case of soluble precipitates.

Let us assume that the rate function u in Eq. (3.5.1.1) depends on (y, t), or $u = u(y, t) \in C^\infty$, and $D(y, t) \in C^\infty$, that is, the properties of the medium (say, its temperature) vary slowly. Let us also assume that the constants a and b have the form $a = \varepsilon^{-1}$ and $b = \varepsilon^{-1} - 1$ (which means that the chemical substances taking part in the reaction and their chemical compounds have a high solubility). We wish to calculate the function $v = v(\varphi)$:

$$\begin{aligned} v &= \left(\frac{b}{a-\varphi}\right)^\lambda = \left(\frac{1-\varepsilon}{1-\varepsilon\varphi}\right)^\lambda \\ &= 1 + \lambda\varepsilon(\varphi - 1) + \lambda\varepsilon^2\left[\frac{\lambda-1}{2}(\varphi-1)^2 \right. \\ &\quad \left. + \varphi(\varphi - 1)\right] + \dots\,. \end{aligned} \tag{3.5.2.1}$$

[5] The boundary conditions for Eq. (3.5.1.1) depend on the type of kinetic equation and will be given below for each case separately.

Substituting this into Eq. (3.5.1.1), we get

$$\varepsilon\left(\lambda+\frac{1}{m}\right)\frac{\partial\varphi}{\partial t}+\lambda\varepsilon u(y,\ t)\frac{\partial\varphi}{\partial y}-\varepsilon^2\lambda\frac{\partial}{\partial y}\left(D(y,\ t)\frac{\partial\varphi}{\partial y}\right)=-\varepsilon^2L_1f(\varphi),$$

$$\varphi\,|_{y=-\psi(t)}=0,\ \ \varphi\,|_{y\to-\infty}\to 1-0,\ \ \frac{\partial\varphi}{\partial y}\Big|_{y\to\infty}\to 0,$$
$$-\infty\leqslant y\leqslant-\psi(t). \tag{3.5.2.2}$$

Here $\psi(t)$ is an unknown function yet to be defined,

$$L_1=\lambda\left(\frac{\partial}{\partial t}+u\frac{\partial}{\partial y}-\varepsilon\frac{\partial}{\partial y}\left(D(y,\ t)\frac{\partial}{\partial y}\right)\right),$$

$$f(\varphi)=\frac{\lambda-1}{2}(\varphi-1)^2+\varphi(\varphi-1).$$

Equation (3.5.2.2) is one with a small nonlinearity. The algorithm of its solution is based on the fact that due to the nonhomogeneity of the principal part of the equation with respect to the orders of the derivatives the nonlinearity generates no resonance terms and can be taken into account by regular perturbation theory methods.

We will seek the asymptotic solution to Eq. (3.5.2.2) in the form

$$\varphi(y,\ t,\ \varepsilon)=[W_0(y,\ t,\ \varepsilon)+\varepsilon W_1(y,\ t,\ \tau)]\,|_{\tau=S/\varepsilon}, \tag{3.5.2.3}$$

where[6]

$$W_i\in B^\alpha,\ \ i=0,\ 1,\ \ S(y,\ t)=\beta(t)(y+\psi(t))+\beta_1(t)(y+\psi(t))^2.$$

It can be easily verified that in this case α must be equal to unity and the W_i satisfy the following conditions:

$$\begin{aligned}&W_0(y,\ t,\ 0)=0,\ \ \lim_{\tau\to-\infty}W_0(y,\ t,\ \tau)=1,\\ &W_1(y,\ t,\ 0)=0,\ \ \lim_{\tau\to-\infty}W_1(y,\ t,\ \tau)=0.\end{aligned} \tag{3.5.2.4}$$

The main result of this section is formulated in

---

[6] Here B^α is the function space introduced in [3.3]. It is said that a function $W(y,\ t,\ \tau)$ belongs to class B^α if

(i) $W(y,\ t,\ \tau)\in C^\infty(R^n\times[0,\ T]\times(R\setminus\{0\}))$;

(ii) for all $(y,\ t)\in R^n\times[0,\ T]$ the function $W(y,\ t,\ \tau)$ is continuous in τ, $W(y,\ t,\ \tau)>0$ for $\tau>0$, and $W(y,\ t,\ \tau)\equiv 0$ for $\tau\leqslant 0$;

(iii) there are functions $\psi_0(y,\ t)\in C^\infty(R^n\times[0,\ T])$ and $\Psi(y,\ t,\ \tau)\in C^\infty\times[R^n\times[0,\ T]\times R_+\setminus\{0\})]$, with $\Psi(y,\ t,\ \tau)=O(\tau^\omega)$ as $\tau\to 0$ and $\omega>0$, such that

$$W(y,\ t,\ \tau)=\tau^\alpha[\psi_0(y,\ t)+\Psi(y,\ t,\ \tau)]\text{ as }\tau\to 0;$$

(iv) $\lim\limits_{\tau\to\infty}W(y,\ t,\ \tau)=V(y,\ t)$, with $V(y,\ t)\in C^\infty$ and the difference $W(y,\ t,\ \tau)-V(y,\ t)$ decreasing faster than any negative power of τ as $\tau\to\infty$;

(v) for every multi-index $\beta\in Z_+^n$ and every $\nu\in Z_+$, the derivatives $\dfrac{\partial^{|\beta|+\nu}}{\partial x^\beta\,\partial t^\nu}W(y,\ t,\ \tau)$ satisfy the conditions stated in assertions (i)-(iv).

Theorem 3.5.2.1 *Let the system of differential equations*

$$\left(1+\frac{1}{m\lambda}\right)\frac{d\psi}{dt}+u(-\psi(t),\ t)-D(-\psi(t),\ t)\beta(t)=0,$$

$$\frac{d\beta}{dt}+2\beta_1\frac{dW}{dt}-\beta\left(1+\frac{1}{m\lambda}\right)^{-1}\Big[\beta\frac{\partial D}{\partial y}\Big|_{y=-\psi(t)}$$

$$+4\beta_1 D(-\psi(t),\ t)-u(-\psi(t),\ t)\frac{\beta_1(t)}{\beta(t)}$$

$$-\frac{\partial u}{\partial y}\Big|_{y=-\psi(t)}\Big]=0,$$

$$2\beta_1 D(-\psi(t),t)+\beta\frac{\partial D}{\partial y}\Big|_{y=-\psi(t)}$$

$$-\lambda\beta(t)\frac{d\psi}{dt}-\lambda\beta(t)u(-\psi(t),\ t)+\lambda\beta^2 D(-\psi(t),\ t)=0$$

have a smooth solution. Then there exists an asymptotic solution to problem (3.5.2.2) *of the* (3.5.2.3) *type and*

$$W_0=W_0(\tau)=\begin{cases}1-\exp\tau & if\ \ \tau<0,\\ 0 & if\ \ \tau\geqslant 0,\end{cases}$$

$$W_1=W_1(t,\ \tau)=\begin{cases}G(\exp 2\tau-\exp\tau) & if\ \ \tau<0,\\ 0 & if\ \ \tau\geqslant 0,\end{cases}$$

where

$$G\quad(\lambda+1)(2D(-\psi(t),\ t)\beta^2(t)\ \Big[\beta(t)\frac{d\psi(t)}{dt}$$

$$+\beta u(-\psi(t),t)-\beta^2(t)D(-\psi(t),\ t)-1\Big].$$

Proof. Substituting solution (3.5.2.3) into Eq. (3.5.2.2) and nullifying the coefficients of the like powers of ε, we get

$$\varepsilon^0:\quad\Big[\lambda\beta\frac{\partial\psi}{\partial t}+\beta\frac{\partial\psi/\partial t}{m}+\lambda u(-\psi(t),\ t)\beta\Big]\frac{dW_0}{d\tau}$$

$$-D(-\psi(t),\ t)\lambda\beta^2(t)\frac{d^2W_0}{d\tau^2}=0,$$

$$\varepsilon^1:\quad\frac{\partial W_0}{\partial\tau}\left(\lambda+\frac{1}{m}\right)\Big[\frac{d\beta}{dt}+2\beta_1\frac{d\psi}{dt}\Big]\frac{\tau}{\beta}$$

$$-\frac{\partial W_1}{\partial\tau}\left(\lambda+\frac{1}{m}\right)\beta\frac{d\psi}{dt}$$

$$\lambda\frac{\partial W_0}{\partial\tau}\Big[\frac{\partial u}{\partial y}\Big|_{y=-\psi(t)}+\frac{2\beta_1}{\beta}u(-\psi,\ t)\Big]\tau+\lambda\frac{\partial W_1}{\partial\tau}u(-\psi,\ t)\beta$$

$$-\lambda D(-\psi,\ t)\beta^2\frac{\partial^2W_1}{\partial\tau^2}-\lambda\frac{\partial^2W_0}{\partial\tau^2}\Big[4D(-\psi,\ t)\beta_2\tau+\frac{\partial D}{\partial y}\Big|_{y=-\psi}\tau\beta\Big]$$

(3.5.2.5)

$$-\lambda \frac{\partial D}{\partial y}\bigg|_{y=-\psi} \beta \frac{\partial W_0}{\partial \tau} - \lambda D(-\psi, t) \frac{\partial W_0}{\partial \tau} 2\beta_1$$

$$= -\frac{dW_0}{d\tau}\left[\lambda\beta \frac{d\psi}{dt} + \lambda\beta u(-\psi, t)\right][(\lambda-1)(W_0-1)+2W_0-1]$$

$$+\lambda\beta^2 D(-\psi, t)\left(\frac{d^2W_0}{d\tau^2}[(\lambda-1)(W_0-1)+2W_0-1]\right.$$

$$\left.+\left(\frac{dW_0}{d\tau}\right)^2(\lambda+1)\right).$$

Here we have used the following relationships:

$$y+\psi(t) = \frac{\varepsilon\tau}{\beta} + O((\tau\varepsilon)^2)\,|_{\tau=S/\varepsilon},$$

$$\frac{\partial S}{\partial t} = \beta \frac{\partial \psi}{\partial t} + \left(\frac{\partial \beta}{\partial t} + 2\beta_1 \frac{\partial \psi}{dt}\right)\frac{\tau\varepsilon}{\beta} + O((\tau\varepsilon)^2)\bigg|_{\tau=S/\varepsilon},$$

$$\frac{\partial S}{\partial y} = \beta + 2\beta_1 \frac{\tau\varepsilon}{\beta} + O((\tau\varepsilon)^2)\bigg|_{\tau=S/\varepsilon}.$$

The solution to Eq. (3.5.2.5) has the form $W_0 = E + Be^{\gamma\tau}$, where $E = \text{const}$ and

$$\gamma = \frac{(\lambda+1/m)(d\psi/dt)+\lambda u(-\psi(t), t)}{\lambda\beta(t) D(-\psi(t), t)}. \tag{3.5.2.6}$$

This formula and the definition of S_0 imply that we can assume that $\gamma = 1$ without loss of generality (otherwise we can change the definition of $\beta(t)$). The final result is that W_0 is the solution to

$$-\frac{dW_0}{d\tau} + \frac{d^2W_0}{d\tau^2} = 0, \tag{3.5.2.7}$$

with

$$(\lambda+1/m)\frac{\partial\psi(t)}{\partial t} + \lambda u(-\psi(t), t) - \lambda D(-\psi(t), t)\,\beta(t) = 0. \tag{3.5.2.8}$$

Equation (3.5.2.7) yields

$$W_0 = \begin{cases} 1-\exp\tau & \text{if } \tau \leqslant 0, \\ 0 & \text{if } \tau > 0. \end{cases} \tag{3.5.2.9}$$

Nullifying the coefficient of ε^1, we get the following equation

$$\lambda D(-\psi, t)\beta^2 \frac{\partial^2 W_1}{\partial\tau^2} - \frac{\partial W_1}{\partial\tau}\left(\left(\lambda+\frac{1}{m}\right)\beta(t)\frac{d\psi}{dt} + \lambda u(-\psi_1 t)\beta(t)\right) = f, \tag{3.5.2.10}$$

where

$$f = \lambda e^{\tau}\tau\Big[\beta\frac{\partial D}{\partial y}\Big|_{y=-\psi(t)} + 4\beta_1 D(-\psi,\ t)$$
$$-\left(1+\frac{1}{m\lambda}\right)\left(\frac{d\beta}{dt}+2\beta_1\frac{d\psi}{dt}\right)\Big/\beta$$
$$-u(-\psi,\ t)\,2\beta_1/\beta-\frac{\partial u}{\partial y}\Big|_{y=-\psi(t)}\Big]$$
$$+\lambda e^{\tau}\Big[2\beta_1 D(-\psi,\ t)+\beta\frac{\partial D}{\partial y}\Big|_{y=-\psi}-\lambda\beta\frac{d\psi}{dt}-\lambda\beta u(-\psi,\ t)$$
$$+\lambda\beta^2 D(-\psi,\ t)\Big]+\lambda e^{2\tau}\Big[\Big(\beta\frac{d\psi}{dt}-\beta u(-\psi,\ t)$$
$$-\beta^2 D(-\psi,\ t)-1\Big)(\lambda+1)\Big]. \qquad (3.5.2.11)$$

By virtue of (3.5.2.8), Eq. (3.5.2.10) can be rewritten thus:

$$\lambda D(-\psi,\ t)\,\beta^2\left(\frac{\partial^2 W_1}{\partial\tau^2}-\frac{dW_1}{\partial\tau}\right)=f.$$

From this it follows that if the expressions inside the first two

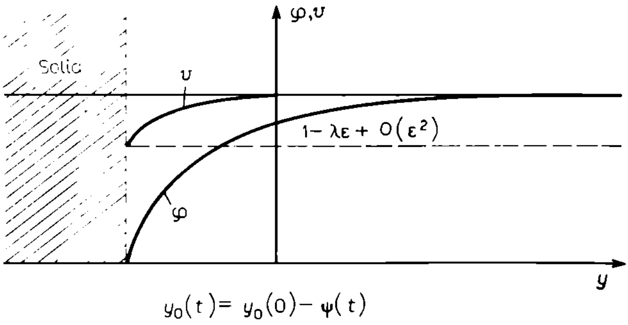


Fig. 3.19

pairs of the square brackets in (3.5.2.11) vanish, that is, if

$$\beta\frac{\partial D}{\partial y}\Big|_{y=-\psi}+4\beta_1 D(-\psi,\ t)-\left(1+\frac{1}{m\lambda}\right)\left(\frac{d\beta}{dt}+2\beta_1\frac{d\psi}{dt}\right)\Big/\beta$$
$$-u(-\psi,\ t)\,2\beta_1/\beta-\frac{\partial u}{\partial y}\Big|_{y=-\psi}=0, \qquad (3.5.2.12)$$

$$2\beta_1 D(-\psi,\ t)+\beta\frac{\partial D}{\partial y}\Big|_{y=-\psi}-\beta\frac{\partial\psi}{\partial t}-\beta u(-\psi,\ t)+\beta^2 D(-\psi,\ t)=0,$$

then the particular solution $\hat{W}_1$ to Eq. (3.5.2.10) has the form

$$\hat{W}_1 = G \exp 2\tau,\ G = (2D(-\psi,\ t)\beta^2)^{-1}(\beta\psi + \beta u(-\psi,\ t) - \beta^2 D(-\psi,\ t) - 1).$$

The general solution to Eq. (3.5.2.10) satisfying the boundary conditions (3.5.2.4) has, therefore, the form

$$W_1 = G \exp 2\tau - G \exp \tau$$

(note that $G \exp \tau$ appears as a solution to the homogeneous counterpart of Eq. (3.5.2.10)). The relationships in (3.5.2.12) can easily be transformed into those that are present in the theorem, and this completes the proof of the theorem.

The graphs of concentrations $\varphi\ (y,\ t,\ \varepsilon)$ and $v\ (y,\ t,\ \varepsilon)$ are depicted in Figure 3.19, where $y_0\ (0) = \text{const} = -\psi\ (0)$ specifies the initial position of the phase boundary.

3.5.3 An Asymptotic Solution to the Kinetic Equation for Nonequilibrium Molecular Processes that Allows for External Diffusion Effects

In this section we study a mathematical model of nonequilibrium precipitation, assuming that the solubility of the substances participating in the process is high: $a = \varepsilon^{-1}$ and $b = a - 1 = \varepsilon^{-1} - 1$.

Let us take the system of equations

$$\begin{aligned} &\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial y} + \frac{\varepsilon}{m}\frac{\partial \varphi}{\partial t} - \varepsilon\frac{\partial}{\partial y}\left(D\frac{\partial v}{dy}\right) = 0, \\ &\frac{\partial \varphi}{\partial t} = \mu_1\left[v - \left(\frac{b}{a-\varphi}\right)^{\lambda}\right]. \end{aligned} \tag{3.5.3.1}$$

Here the boundary conditions for φ are the same as in Eq. (3.5.2.2), and the boundary conditions for v are

$$v\,|_{y\to-\infty} \to 1,\ v\,|_{y=-\psi(t)} = 1 - \lambda\varepsilon - \varepsilon^2\frac{\lambda(\lambda-1)}{2}.$$

Let us assume that $\mu_1 = \varepsilon^{-2}\mu$, with $\mu = \text{const}$. This means that the rate of the internal diffusion of the substance to the sorbent particles is fairly high. The solution to the system of equations (3.5.3.1) will be sought in the form

$$\begin{aligned} &\varphi = [W_0(y,\ t,\ \tau) + \varepsilon W_1(y,\ t,\ \tau)]\,|_{\tau=S/\varepsilon}, \\ &v = [Z_0(y,\ t,\ \tau) + \varepsilon Z_1(y,\ t,\ \tau) \\ &\quad + \varepsilon^2 Z_2(y,\ t,\ \tau)]\,|_{\tau=S/\varepsilon}, \end{aligned} \tag{3.5.3.2}$$

where $S\ (y,\ t) = \beta\ (t)\ (y + \psi\ (t)) + \beta_1\ (t)\ (y + \psi\ (t))^2$, Z_i, $W_i \in B^1$, and, as $\tau \to -\infty$, $W_1 = O\ (Z_1)$ and $Z_2 = O\ (Z_1)$. The conditions

similar to (3.5.2.4) in this case are

$$\begin{aligned} &W_0(y,\ t,\ 0)=0, \qquad \lim_{\tau\to-\infty} W_0(y,\ t,\ \tau)=1,\\ &W_1(y,\ t,\ 0)=0, \qquad \lim_{\tau\to-\infty} W_1(y,\ t,\ \tau)=0, \end{aligned} \tag{3.5.3.3}$$

$$\begin{aligned} &Z_0(y,\ t,\ 0)=1, \qquad \lim_{\tau\to-\infty} Z_0(y,\ t,\ \tau)=1,\\ &Z_1(y,\ t,\ 0)=-\lambda, \quad \lim_{\tau\to-\infty} Z_i(y,t,\tau)=0,\ i\geqslant 1.\\ &Z_2(y,\ t,\ 0)=\lambda(\lambda-1)/2, \end{aligned} \tag{3.5.3.4}$$

and the conditions for the Z_i at $\varepsilon = 0$ follow from the expansion

$$v\,|_{\tau=0}=(b/a)^{\lambda}=1-\lambda\varepsilon+\varepsilon^2\lambda(\lambda-1)/2+O(\varepsilon^3).$$

The main result of this section is formulated in the following

Theorem 3.5.3.1 *Under the above assumptions, there exists an asymptotic solution to* (3.5.3.1) *satisfying conditions* (3.5.3.3) *and* (3.5.3.4) *if the system of equations* (3.5.3.26) *has smooth solutions and if* $(d\psi/dt)^{-1}(d\psi/dt+u(-\psi,\ t))<-(\lambda m)^{-1}$. *Here* $Z_0\equiv 1$ *and*

$$\begin{pmatrix} W_0\\ Z_1\end{pmatrix}=-\left(\lambda+\frac{(\gamma_1-a-\lambda b)(\gamma_2-a)}{b}\right)\begin{pmatrix}\dfrac{b}{\gamma_1-a}\\ 1\end{pmatrix}\exp\gamma_1\tau$$
$$+\frac{(\gamma_1-a-\lambda b)(\gamma_2-a)}{b}\begin{pmatrix}\dfrac{b}{\gamma_2-a}\\ 1\end{pmatrix}\exp\gamma_2\tau+\begin{pmatrix}1\\0\end{pmatrix}, \tag{3.5.3.5}$$

where

$$a=-\lambda\mu\left(\beta\frac{d\psi}{dt}\right)^{-1},\ b=\mu\left(\beta\frac{d\psi}{dt}\right)^{-1},$$
$$\gamma_{1,2}=(a+d\pm\sqrt{(a-d)^2+4cb})/2,$$
$$c=\frac{d\psi}{dt}(mD(-\psi,\ t)\beta),$$
$$d=\left(\frac{d\psi}{dt}+u(-\psi,\ t)\right)(D\beta)^{-1}.$$

The functions W_1 *and* Z_2 *will be defined later in formulas* (3.5.3.24), *and the functions* β, ψ, *and* β_1 *can be found by solving the system of equations* (3.5.3.26).

Proof. Substituting into (3.5.3.1) the expansion $[b/(a-\varphi)]^{\lambda}=1+\lambda\varepsilon(\varphi-1)+O(\varepsilon^2)$, we can rewrite system (3.5.3.1) as

follows:

$$\varepsilon \frac{\partial v}{\partial t} + u\varepsilon \frac{\partial v}{\partial y} + \frac{\varepsilon^2}{m_1} \frac{\partial \varphi}{\partial t} - \varepsilon^2 \frac{\partial}{\partial y}\left(D_1 \frac{\partial v}{\partial y}\right) = 0,$$

$$\varepsilon \frac{\partial \varphi}{\partial t} = \frac{1}{\varepsilon} \mu_0 (v-1) - \mu\lambda(\varphi - 1) - \varepsilon\lambda[((\lambda-1)/2)(\varphi-1)^2 + \varphi(\varphi-1)] + O(\varepsilon^2). \qquad (3.5.3.6)$$

The function $S(y, t)$ has the form

$$S(y, t) = \beta(t)(y + \psi(t)) + \beta_1(t)(y + \psi(t))^2.$$

It can easily be verified that

$$\frac{\partial S}{\partial t} = \beta \frac{d\psi}{dt} + \left(\frac{d\beta}{dt} + 2\beta_1 \frac{d\psi}{dt}\right)\left(\frac{\tau\varepsilon}{\beta} - \frac{\beta_1(\tau\varepsilon)^2}{\beta^3}\right) + \frac{d\beta_1}{dt}\frac{(\varepsilon\tau)^2}{\beta^2} + O((\varepsilon\tau)^3)\Big|_{\tau=S/\varepsilon},$$

$$\frac{\partial S}{\partial x} = \beta + 2\beta_1\left(\frac{\varepsilon\tau}{\beta} - \frac{\beta_1(\tau\varepsilon)^2}{\beta^3}\right) + O((\varepsilon\tau)^3)\Big|_{\tau=S/\varepsilon}.$$

Substituting into (3.5.3.5) and (3.5.3.6) the formulas for the solutions (3.5.3.2) given at the beginning of this section, nullifying the coefficients of the higher powers of ε, and allowing for the estimates of the derivatives $\partial S/\partial t$ and $\partial S/\partial x$, we get

$$\varepsilon^0: \left(\frac{\partial \psi}{\partial t} + u(-\psi, t)\right)\frac{dZ_0}{d\tau} - D(-\psi(t), t)\beta \frac{d^2 Z_0}{d\tau^2} = 0,$$

$$\varepsilon^{-1}: Z_0 - 1 = 0. \qquad (3.5.3.6')$$

Obviously, the only solution to the last system of equations is the function

$$Z_0 \equiv 1. \qquad (3.5.3.7)$$

Conditions (3.5.3.4) are also met. Substituting the solutions (3.5.3.2) into system (3.5.3.1) and nullifying the coefficients of the first and zero powers of ε, we get

$$\varepsilon^1: \frac{\partial Z_1}{\partial \tau}\beta\frac{\partial \psi}{\partial t} + u(-\psi, t)\beta\frac{\partial Z_1}{\partial \tau} + \frac{\beta}{m}\frac{\partial \psi}{\partial t}\frac{\partial W_0}{\partial \tau} - D(-\psi, t)\beta^2 \frac{\partial^2 Z_1}{\partial \tau^2} = 0, \qquad (3.5.3.8)$$

$$\varepsilon^0: \frac{\partial W_0}{\partial \tau}\beta\frac{\partial \psi}{\partial t} = \mu Z_1 - \lambda\mu(W_0 - 1).$$

The boundary conditions for this system of equations are as follows:

$$W_0|_{\tau=0} = 0, \quad \lim_{\tau\to-\infty} W_0 = 1,$$

$$Z_1|_{\tau=0} = -\lambda, \quad \lim_{\tau\to-\infty} Z_1 = 0. \qquad (3.5.3.9)$$

Integrating the first equation in (3.5.3.8), we get, in view of the boundary conditions as $\tau \to -\infty$,

$$\left(\beta \frac{\partial\psi}{\partial t} + u(-\psi, t)\beta\right) Z_1 - D_1(-\psi, t)\beta^2 \frac{\partial Z_1}{\partial\tau}$$
$$= -\frac{\beta}{m_1}\frac{\partial\psi}{\partial t}(W_0 - 1). \qquad (3.5.3.10)$$

Thus, for determining W_0 and Z_1 we have the following system of ordinary differential equations:

$$\begin{aligned} \frac{\partial W_0}{\partial\tau} - \frac{\mu}{\beta(d\psi/dt)} Z_1 + \frac{\lambda\mu}{\beta(d\psi/dt)} W_0 &= \frac{\lambda\mu_0}{\beta(d\psi/dt)}, \\ \frac{\partial Z_1}{\partial\tau} - \frac{d\psi/dt + u}{D\beta} Z_1 - \frac{d\psi/dt}{mD\beta} W_0 &= -\frac{d\psi/dt}{mD\beta}. \end{aligned} \qquad (3.5.3.11)$$

We denote by A a matrix of the form $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, with

$$\begin{aligned} a &= -\frac{\lambda\mu}{\beta(d\psi/dt)}, & b &= \frac{\mu}{\beta(d\psi/dt)}, \\ c &= \frac{d\psi/dt}{mD\beta}, & d &= (d\psi/dt + u)/(D\beta), \end{aligned} \qquad (3.5.3.12)$$

and $D = D(-\psi, t)$ and $u = u(-\psi, t)$. Then the general solution to system (3.5.3.11) is

$$\begin{pmatrix} W_0 \\ Z_1 \end{pmatrix} = C_1 \mathbf{l}_1 \exp \gamma_1\tau + C_2 \mathbf{l}_2 \exp \gamma_2\tau + A^{-1}\begin{pmatrix} -\lambda\mu_0/(\beta d\psi/dt) \\ (d\psi/dt)/(D\beta m) \end{pmatrix}, \qquad (3.5.3.13)$$

where C_1 and C_2 are arbitrary constants (i.e. functions of variable t), γ_1 and γ_2 are the eigenvalues of matrix A, and $\mathbf{l}_1$ and $\mathbf{l}_2$ are the corresponding eigenvectors of this matrix (we will shortly show that γ_1 can be assumed unequal to γ_2 without any loss of generality).

For solution (3.5.3.13) to satisfy the boundary conditions, at least one of the eigenvalues must be positive. Obviously, the formula for the eigenvalues is

$$\gamma_{1,2} = \frac{a + d \pm \sqrt{(a-d)^2 + 4bc}}{2}, \qquad (3.5.3.14)$$

or, in greater detail,

$$\gamma_{1,2} = \frac{1}{2}\left\{\frac{d\psi/dt + u(-\psi, t)}{D\beta} - \frac{\mu\lambda}{\beta(d\psi/dt)}\right.$$
$$\left.\pm \sqrt{[\mu\lambda/(\beta\, d\psi/dt) + (d\psi/dt + u)/D\beta]^2 + 4\mu/(mD\beta)}\right\}.$$

The product cb is equal to $\mu/(mD\beta^2)$ and, hence, $\gamma_1 \neq \gamma_2$ for $\mu > 0$.

Let us assume that the eigenvalues of matrix A are positive. This condition is equivalent to the following inequalities:

$$a + d > 0, \quad ad - bc > 0,$$

or, allowing for (3.5.3.12),

$$-\lambda\mu/(d\psi/dt) < (d\psi/dt + u)/D,$$
$$-(d\psi/dt + u)/(d\psi/dt) > 1/m\lambda.$$

The second condition implies that a solution that is bounded as $\tau \to \infty$ exists only if

$$\frac{d\psi}{dt}\left(\frac{d\psi}{dt} + u\right) < 0.$$

(Note that this condition is met for the exact solution given in Section 3.1.3, where $d\psi/dt = -C = -um/(m+1)$.)

The final necessary condition that a solution of (3.5.3.2) exists is formulated in the form of two inequalities:

$$-\frac{u}{1+1/(\lambda m)} < \frac{d\psi}{dt} < 0, \quad \frac{d\psi}{dt} + u > 0.$$

In this case the solution that is bounded as $\tau \to -\infty$ has the form

$$\begin{pmatrix} W_0 \\ Z_1 \end{pmatrix} = C_1 \begin{pmatrix} \dfrac{b}{\gamma_1 - a} \\ 1 \end{pmatrix} \exp \gamma_1\tau + C_2 \begin{pmatrix} \dfrac{b}{\gamma_2 - a} \\ 1 \end{pmatrix} \exp \gamma_2\tau + \begin{pmatrix} l \\ f \end{pmatrix}. \tag{3.5.3.15}$$

The boundary conditions yield the following system of equations:

$$(\tau = 0) \quad f + C_1 + C_2 = -\lambda,$$
$$l + C_1 b/(\gamma_1 - a) + C_2 b/(\gamma_2 - a) = 0, \tag{3.5.3.16}$$
$$(\tau \to -\infty) \quad l \equiv 1, \ f = 0. \tag{3.5.3.17}$$

Equations (3.5.3.12) yield

$$C_1 = -\lambda - C_2, \quad C_2 = (\gamma_1 - a - \lambda b)(\gamma_2 - a)/b, \tag{3.5.3.18}$$

It can easily be verified that Eq. (3.5.3.17) is satisfied identically.

Let us consider the system of equations for the next approximation:

$$\beta \frac{d\psi}{dt}\frac{\partial W_1}{\partial \tau} - \mu Z_2 + \mu\lambda W_1 + \lambda\left(\frac{\lambda+1}{2} W_0^2 - \lambda W_0 + \frac{\lambda - 1}{2}\right) = f_1, \tag{3.5.3.19}$$

$$\frac{\partial Z_2}{\partial \tau}\left(\beta \frac{d\psi}{dt} + u(-\psi, t)\beta\right) + \frac{\beta}{m_1}\frac{d\psi}{dt}\frac{\partial W_1}{\partial \tau} - D(-\psi, t)\beta^2 \frac{\partial^2 Z_2}{\partial \tau^2} = f_2,$$

with

$$f_1 = -\frac{\partial W_0}{\partial t} - \frac{d\beta}{dt}\frac{\partial W_0}{\partial \tau} - 2\frac{d\psi}{dt}\tau\beta_1\frac{\partial W_0}{\partial \tau},$$

$$\begin{aligned} f_2 = &-\frac{\partial Z_1}{\partial t} + \frac{1}{\beta}\frac{d\beta}{dt}\tau\frac{\partial Z_1}{\partial \tau} + 2\frac{d\psi}{dt}\left(\tau\beta_1\frac{\partial Z_1}{\partial \tau}\right)\frac{1}{\beta} \\ &+ \frac{\partial u}{\partial y}\bigg|_{y=-\psi}\tau\frac{\partial Z_1}{\partial \tau} + u(-\psi, t)\, 2\beta_1\tau\frac{\partial Z_1}{\partial \tau}\frac{1}{\beta} \\ &+ \frac{1}{m}\frac{\partial W_0}{\partial t} + \frac{\beta}{m}\frac{\partial W_0}{\partial \tau}\frac{d\psi}{dt} - \beta\frac{\partial D}{\partial y}\bigg|_{y=-\psi}\frac{\partial Z_1}{\partial \tau} \\ &- 2D(-\psi, t)\beta_1\frac{\partial Z_1}{\partial \tau} \\ &- D(-\psi, t)\, 4\tau\beta_1\frac{\partial^2 Z_1}{\partial \tau^2} - \beta\frac{\partial D}{\partial y}\bigg|_{y=-\psi}\tau\frac{\partial^2 Z_1}{\partial \tau^2}. \end{aligned}$$

The system of equations (3.5.3.19) can be solved in a manner similar to that used in solving Eq. (3.5.2.10).

Let us transform the expressions for f_1 and f_2 using (3.5.3.6′) and (3.5.3.7):

$$\begin{aligned} f_1 = &\; G_1\exp\{\gamma_1\tau\} + G_2\exp\{\gamma_2\tau\} + K_1\tau\exp\{\gamma_1\tau\} + K_2\tau\exp\{\gamma_2\tau\} \\ &+ B_0\exp\{(\gamma_1+\gamma_2)\tau\} + B_1\exp\{2\gamma_1\tau\} + B_2\exp\{2\gamma_2\tau\}, \end{aligned} \tag{3.5.3.20}$$

$$\begin{aligned} f_2 = &-[N_1(t)\exp\{\gamma_1\tau\} + N_2\exp\{\gamma_2\tau\} + \Phi_1\tau\exp\{\gamma_1\tau\} \\ &+ \Phi_2\tau\exp\{\gamma_2\tau\}], \end{aligned}$$

where

$$G_i = -C_i l_{i1}\lambda - \frac{\partial(C_i l_{i1})}{\partial t} - C_i l_{i1}\gamma_i\frac{d\beta}{dt},$$

$$K_i = -C_i l_{i1}\frac{\partial\gamma_i}{\partial t}\beta_1 C_i l_{i1}\gamma_i,$$

$$B_i = -\lambda(\lambda+1)C_i^2 l_{i1}^2/2, \quad i = 1, 2,$$

$$B_0 = -\lambda(\lambda+1)C_1C_2 l_{11}l_{12},$$

$$\begin{aligned} N_1 = \Big\{&\frac{d}{dt}\left(C_1 l_{12} + \frac{C_1 l_{11}}{m}\right) + \Big(\frac{\beta}{m}\frac{d\psi}{dt}l_{11} \\ &- \beta\frac{\partial D}{\partial y}\bigg|_{y=-\psi}l_{12} - 2D(-\psi, t)\beta_1 l_{11}\Big)C_1\gamma_1\Big\}, \end{aligned}$$

$$\begin{aligned} N_2 = \Big\{&\frac{d}{dt}(C_2 l_{22} + C_2 l_{21}/m) + \Big(\frac{\beta}{m}\frac{d\psi}{dt}l_{22} \\ &- \beta\frac{\partial D}{\partial y}\bigg|_{y=-\psi}l_{21} - 2D(-\psi, t)\beta_1 l_{22}\Big)C_2\gamma_2\Big\}, \end{aligned}$$

$$\begin{aligned} \Phi_1 = \Big(&\frac{\partial\gamma_1}{\partial t} + \frac{1}{\beta}\frac{d\beta}{dt}\gamma_1 + 2\frac{d\psi}{dt}\beta_1\gamma_1 \\ &+ \frac{\partial u}{\partial y}\bigg|_{y=-\psi}\gamma_1 - 2u(-\psi, t)\beta_1\gamma_1/\beta - 4D(-\psi, t)\beta_1\gamma_1^2 \end{aligned}$$

$$
\begin{aligned}
&-\beta \left.\frac{\partial D}{\partial y}\right|_{y=-\psi} \gamma_1^2\Big) C_1 l_{12} + \frac{1}{m} C_1 \gamma_1 \frac{\partial \gamma_1}{\partial t}, \\
\Phi_2 &= \Big(\frac{\partial \gamma_2}{\partial t} + \frac{1}{\beta} \frac{d\beta}{dt} \gamma_2 + 2 \frac{d\psi}{dt} \beta_1 \gamma_2 \\
&+ \left.\frac{\partial u}{\partial y}\right|_{y=-\psi} \gamma_2 + 2u(-\psi, t)\beta_1 \gamma_2/\beta \\
&- 4D(-\psi, t)\beta_1 \gamma_2^2 - \beta \left.\frac{\partial D}{\partial y}\right|_{y=-\psi} \gamma_2^2\Big) C_2 l_{22} \\
&+ \frac{1}{m} C_2 l_{21} \frac{\partial \gamma_2}{\partial t}.
\end{aligned}
$$

Here

$$
\begin{aligned}
&l_{11} = b/(\gamma_1 - a), \ l_{12} = 1, \ l_{21} = b/(\gamma_2 - a), \ l_{22} = 1, \\
&C_1 = -\lambda - (\gamma_1 - a - \lambda b)(\gamma_2 - a)/b, \\
&C_2 = (\gamma_1 - a - \lambda b)(\gamma_2 - a)/b.
\end{aligned}
$$

Integrating the second equation in (3.5.3.19) with respect to τ from $-\infty$ to τ, we reduce system (3.5.3.19) to a system of first-order equations:

$$
\begin{aligned}
&\frac{\partial W_1}{\partial \tau} - Z_1 \mu \Big/ \left(\beta \frac{d\psi}{dt}\right) + W_1 \mu\lambda \Big/ \left(\beta \frac{d\psi}{dt}\right) = f_1, \\
&\frac{\partial Z_2}{\partial \tau} - \left(\frac{d\psi}{dt} + u\right) Z_2/(D\beta) \\
&\qquad - W_1 \frac{d\psi}{dt} \Big/ (mD\beta) = \tilde{f}_2,
\end{aligned}
\tag{3.5.3.21}
$$

where f_1 is defined in (3.5.3.20) and

$$
\begin{aligned}
\tilde{f}_2 &= \frac{N_1}{\gamma_1} \exp(\gamma_1 \tau) + \frac{N_2}{\gamma_2} \exp(\gamma_2 \tau) + \frac{\Phi_1}{\gamma_1} (\tau - 1) \exp(\gamma_1 \tau) \\
&+ \frac{\Phi_2}{\gamma_2} (\tau - 1) \exp(\gamma_2 \tau).
\end{aligned}
$$

Note that in view of the above remark, $\gamma_1 \neq \gamma_2$. Precisely, in view of (3.5.3.14) we have $\gamma_1 > \gamma_2$, which implies that, for $-\infty < \tau < 0$,

$$
\exp(\gamma_1 \tau) + \tau^k \exp(\gamma_2 \tau) = O(\exp(\gamma_1 \tau)). \tag{3.5.3.22}
$$

This estimate shows that the terms in the solution to system (3.5.3.21) that have the form $\tau^k \exp(\gamma_2 \tau)$ do not "worsen" the general estimate of the solution in the presence of terms of the $\exp(\gamma_1 \tau)$.

Hence the solution to (3.5.3.21) that allows for the estimate $O(\exp(\gamma_1 \tau))$ exists if

$$
\begin{aligned}
&\Phi_1 = 0, \ K_1 = 0, \\
&\left\langle \begin{pmatrix} G_1 \\ N_1/\gamma_1 \end{pmatrix}, \mathbf{l}_1^* \right\rangle = 0,
\end{aligned}
\tag{3.5.3.23}
$$

where $\mathbf{l}_1^*$ is the eigenvector of matrix A^* that corresponds to the eigenvalue γ_1. The solution to system (3.5.3.21) has the form

$$\begin{pmatrix} W_1 \\ Z_2 \end{pmatrix} = C_3 \begin{pmatrix} b/(\gamma_1 - a) \\ 1 \end{pmatrix} \exp(\gamma_1 \tau) + C_4 \begin{pmatrix} b/(\gamma_2 - a) \\ 1 \end{pmatrix} \exp(\gamma_2 \tau)$$
$$+ \frac{1}{ad - cb} \left[\begin{pmatrix} dK_2 - b\Phi_2/\gamma_2 \\ -cK_2 + a\Phi_2/\gamma_2 \end{pmatrix} \tau \exp(\gamma_2 \tau) \right.$$
$$+ B_0 \begin{pmatrix} d \\ -c \end{pmatrix} \exp((\gamma_1 + \gamma_2)\tau) + B_1 \begin{pmatrix} d \\ -c \end{pmatrix} \exp(2\gamma_1 \tau)$$
$$\left. + B_2 \begin{pmatrix} d \\ -c \end{pmatrix} \exp(2\gamma_2 \tau) \right]$$
$$+ (\gamma_1 I - A)^{-1} \begin{pmatrix} G_1 \\ N_1/\gamma_1 \end{pmatrix} \exp(\gamma_1 \tau). \qquad (3.5.3.24)$$

Note that vector $(\gamma_1 I - A)^{-1} \begin{pmatrix} G_1 \\ N_1/\gamma_1 \end{pmatrix}$ is well-defined thanks to conditions (3.5.3.23) (to within the elements of the kernel of matrix $\gamma_1 I - A$); we will select a fixed value for this vector.

The boundary conditions for the functions W_1 and Z_2 have the form

$$\begin{aligned} &W_1|_{\tau=0} = 0, \quad W_1|_{\tau \to -\infty} \to 0, \\ &Z_2|_{\tau=0} = 0, \quad Z_2|_{\tau \to -\infty} \to 0. \end{aligned} \qquad (3.5.3.25)$$

Substituting solution (3.5.3.24) into (3.5.3.25), we get a system of equations for finding C_3 and C_4:

$$C_3 b/(\gamma_1 - a) + C_4 b/(\gamma_2 - a) + B_0 d + B_1 d + B_2 d$$
$$+ \frac{b}{(\gamma_2 - a)(\gamma_1 - \gamma_2)\langle \mathbf{l}_1, \mathbf{l}_2^* \rangle} \left\langle \begin{pmatrix} G_1 \\ N_1/\gamma_1 \end{pmatrix}, \mathbf{l}_2^* \right\rangle = 0,$$
$$C_3 + C_4 - B_0 c - B_1 c + \frac{1}{(\gamma_1 - \gamma_2)\langle \mathbf{l}_2, \mathbf{l}_1^* \rangle} \left\langle \begin{pmatrix} G_1 \\ N_1/\gamma_1 \end{pmatrix}, \mathbf{l}_2^* \right\rangle = 0.$$

Solving the system, we get (see p. 310)

$$C_4 = -\frac{(B_0 + B_1 + B_2 + B_4)(d\gamma_1 + da + bc)(\gamma_2 - a)}{b(\gamma_1 - \gamma_2)},$$
$$C_3 = -C_4 + (B_0 + B_1 + B_2 + B_5)c,$$

where

$$B_4 = \frac{b}{d(\gamma_2 - a)(\gamma_1 - \gamma_2)\langle \mathbf{l}_2, \mathbf{l}_2^* \rangle} \left\langle \begin{pmatrix} G_1 \\ N_1/\gamma_1 \end{pmatrix}, \mathbf{l}_2^* \right\rangle,$$
$$B_5 = \frac{1}{c(\gamma_1 - \gamma_2)\langle \mathbf{l}_2, \mathbf{l}_2^* \rangle} \left\langle \begin{pmatrix} G_1 \\ N_1/\gamma_1 \end{pmatrix}, \mathbf{l}_2^* \right\rangle,$$

and $\mathbf{l}_2^*$ is the eigenvector of A^* corresponding to eigenvalue γ_2.

Equations (3.5.3.23) are the equations for finding the unknown functions $\psi(t)$, $\beta(t)$, and $\beta_1(t)$. Let us rewrite these equations by employing the notations introduced above:

$$\begin{aligned}
&\frac{d\gamma_1}{dt}+2\frac{d\psi}{dt}\beta_1\gamma_1=0,\\
&\frac{1}{\beta}\frac{d\beta}{dt}\gamma_1+\left.\frac{\partial u}{\partial y}\right|_{y=-\psi}\gamma_1+2u(-\psi,\ t)\beta_1\gamma_1/\beta\\
&\qquad -4D(-\psi,\ t)\beta_1\gamma_1^2\\
&\qquad -\gamma_1^2\beta\left.\frac{\partial D}{\partial y}\right|_{y=-\psi}+\frac{b}{m(\gamma_1-a)}=0,\\
&\frac{\gamma_1-a}{\gamma_1}\frac{d}{dt}(C_1+C_1l_{11})+C_1\left(\frac{\beta}{m}\frac{d\psi}{dt}l_{11}-\beta\left.\frac{\partial D}{\partial y}\right|_{y=-\psi}\right.\\
&\left.-2D(-\psi,\ t)\beta_1l_{11}\right)-C_1l_{11}\lambda-\frac{d}{dt}(C_1l_{11})-C_1l_{11}\gamma_1\frac{d\beta}{dt}=0.
\end{aligned}\tag{3.5.3.26}$$

When the eigenvalues (3.5.3.14) of matrix A have different signs, the solution to (3.5.3.11) that is bounded as $\tau\to-\infty$ has the form (for $\gamma_1>0$)

$$\begin{pmatrix}W_0\\ Z_1\end{pmatrix}=\begin{pmatrix}1\\ 0\end{pmatrix}+\begin{pmatrix}-1\\ -\lambda\end{pmatrix}\exp(\gamma_1\tau).$$

Another necessary condition here is

$$\gamma_1-a-\lambda b=0. \tag{3.5.3.27}$$

Other relationships linking ψ, β, and β_1 that are similar to (3.5.3.26) can be obtained from the existence condition on $\begin{pmatrix}W_1\\ Z_2\end{pmatrix}$. The right-hand sides of the system of equations (3.5.3.19) are determined by formulas (3.5.3.20), where we must put

$$C_2=0,\quad l_{11}=-1/\lambda,\quad l_{12}=-1,\quad C_1=-\lambda. \tag{3.5.3.28}$$

A necessary condition for the existence of a solution to the system of equations (3.5.3.11) with an estimate $O(\exp(\gamma_1\tau))$ as $\tau\to-\infty$ is (3.5.3.23) (where we have allowed for (3.5.3.28)). It can easily be verified that the system of equations (3.5.3.23) augmented with Eq. (3.5.3.27) is overdetermined (four equations and three unknowns $\psi(t)$, $\beta(t)$, and $\beta_1(t)$). Therefore, when the eigenvalues of matrix A have different signs, the initial problem has no asymptotic solution of the (3.5.3.2) type.

3.5.4 Equilibrium Precipitation (Coprecipitation) Accompanied by Other Intensive Chemical Reactions

In this section we build an asymptotic solution for a system describing precipitation in the case where the source-sink function has zeros.

In this case Eq. (3.5.2.1) is replaced by

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial y}+\frac{\varepsilon}{m}\frac{\partial\varphi}{\partial t}=\varepsilon D\frac{\partial^2 v}{\partial y^2}-F(v,\ y,\ t). \tag{3.5.4.1}$$

This equation is augmented with the isotherm equation of precipitation, $\varphi=\varphi(v)$.

Let us consider the case where $F(v, y, t)$ satisfies the conditions usually imposed on the source-sink function in the KPP equation, which was extensively discussed in Sections 3.2 and 3.6. Thus, we have the problem

$$\begin{gathered}\varepsilon\left(\frac{\partial v}{\partial t}+u(y,\ t)\frac{\partial v}{\partial y}\right)+\frac{\varepsilon^2}{m}\frac{\partial\varphi}{\partial t}\\ -\varepsilon^2\frac{\partial}{\partial y}\left(D(y,\ t)\frac{\partial v}{\partial y}\right)-\frac{F(v,\ y,\ t)}{\varepsilon}=0,\\ \varphi=a-b/v^{1/\lambda},\ F(1)=0,\ F(A)=0,\\ 0<A<1,\ 0<\varepsilon<1,\end{gathered} \tag{3.5.4.2}$$

where F can be equal, say, to $d(y, t)(A(y, t)-v)(v-1)$, $d>0$ with u, D, and A positive and continuously differentiable functions.

The boundary conditions have the form

$$\begin{gathered}v\Big|_{\substack{\varphi=0\\ y=y_*(t)}}=v_0(t),\ \ \varphi|_{y=y_*(t)}=0,\\ v|_{y\to-\infty}=1,\ \ \varphi|_{y\to-\infty}=1.\end{gathered} \tag{3.5.4.3}$$

We divide the differential equation in (3.5.4.2) by ε and transform it to

$$\begin{gathered}\frac{\varepsilon^2 b}{m\lambda}\frac{\partial v}{\partial t}+\varepsilon v^{(\lambda+1)/\lambda}\left[\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial y}\right.\\ \left.-\varepsilon\frac{\partial}{\partial y}\left(D\frac{\partial v}{\partial y}\right)+\varepsilon^{-1}F(v,\ y,\ t)\right]=0.\end{gathered} \tag{3.5.4.4}$$

We seek the solution v in the form

$$v(y,\ t,\ \varepsilon)=[W_0(y,\ t,\ \tau)+\varepsilon W_1(y,\ t,\ \tau)]|_{\tau=S/\varepsilon}, \tag{3.5.4.5}$$

with $S(x,\ t,\ \varepsilon)=\beta(t)(x-\psi)+\beta_1(x,\ t)(x-\psi)^2+\varepsilon S_1(x,\ t)$. The asymptotic solution to the given problem is given in the following

Theorem 3.5.4.1 *Let the system of ordinary differential equations*

$$\beta\left(\frac{d\psi}{dt}+u(-\psi,\ t)\right)=(1-A(-\psi,\ t))\,b_0 d(-\psi,\ t),$$
$$D_1(-\psi,\ t)\,\beta^2=(1-A(-\psi,\ t))d(-\psi,\ t),\ b_0=\mathrm{const}>2,$$

have smooth solutions $\beta=\beta(t)$ *and* $\psi=\psi(t)$. *Then there exists an asymptotic solution to* (3.5.4.4), (3.5.4.3) *of the* (3.5.4.5) *type, with*

$$W_0=W_0(t,\ \tau)=1+(A(-\psi,\ t)-1)\,\chi(\tau),$$

where $\chi(\tau)$ *solves the problem*

$$\begin{gathered} b_0\frac{d\chi}{d\tau}-\frac{d^2\chi}{d\tau^2}-\chi(1-\chi)=0,\\ \chi\,|_{\tau\to-\infty}\to 0,\ \ \chi\,|_{\tau\to+\infty}\to 1. \end{gathered} \tag{3.5.4.6}$$

The function $W_1=W_1(t,\tau)$ *is defined for* $-\infty\leqslant\tau\leqslant 0$ *through the formula*

$$W_1=(A(-\psi,\ t)-1)^{-1}\frac{\partial W_0}{\partial\tau}\int\limits_{-\infty}^{\tau}(\exp(b_0\tau))$$
$$\times\left(\frac{\partial W_0}{\partial\tau}\right)^{-2}\left(\int\limits_0^{\tau}\frac{f(\partial W_0/\partial\tau')}{\exp(b_0\tau')}\,d\tau'\right)d\tau,$$

with b_0 *the speed of the wave, the function* β_1 *will be defined below through formula* (3.5.4.24), *and* S_1 *is the solution to the boundary value problem for Eq.* (3.5.4.23) (*see below*) *satisfying the boundary condition*

$$S_1(x,\ t)\,|_{x=-\psi}=\chi^{-1}\left(\frac{v_0(t)-1}{A(-\psi,\ t)-1}\right), \tag{3.5.4.7}$$

with χ^{-1} *the inverse of* $\chi(\tau)$.

Proof. Substituting (3.5.4.5) into Eq. (3.5.4.4) and nullifying the coefficient of ε^0, we get

$$\frac{\partial W_0}{\partial\tau}\beta\frac{\partial\psi}{\partial t}+u(-\psi,\ t)\ \frac{\partial W_0}{\partial\tau^2}\beta$$
$$-D(-\psi,\ t)\frac{\partial^2 W_0}{\partial\tau^2}\beta^2+F(W_0,\ -\psi(t),\ t)=0. \tag{3.5.4.8}$$

Nullifying the coefficient of ε^1, we get

$$\frac{b}{m\lambda}\frac{\partial W_0}{\partial\tau}\beta\frac{\partial\psi}{\partial t}+W_0^{(\lambda+1)/\lambda}\left\{\frac{\partial W_0}{\partial\tau}\left(\frac{1}{\beta}\frac{\partial\beta_1}{\partial t}+\frac{2\beta_1}{\beta}\frac{\partial\psi}{\partial t}\right)(\tau-S_1)\right.$$
$$+\frac{\partial W_1}{\partial\tau}\beta\frac{\partial\psi}{\partial t}+\beta u(-\psi,\ t)\frac{\partial W_1}{\partial\tau}+u(-\psi,\ t)\ \frac{\partial W_0}{\partial\tau}\frac{2\beta_1}{\beta}\ (\tau-S_1)$$
$$+(\tau-S_1)\frac{\partial u}{\partial y}\bigg|_{y=-\psi}\frac{\partial W_0}{\partial\tau}-D(-\psi,\ t)\ \frac{\partial^2 W_0}{\partial\tau^2}4\beta_1(\tau-S_1)$$

$$-D(-\psi,\ t)\frac{\partial^2 W_1}{\partial\tau^2}\beta^2-\frac{\partial D}{\partial y}\Big|_{y=-\psi}\left(\frac{\tau-S_1}{\beta}\right)\frac{\partial^2 W_0}{\partial\tau^2}$$
$$-\frac{\partial D}{\partial y}\Big|_{y=-\psi}\frac{\partial W_0}{\partial\tau}\beta-D(-\psi,\ t)\frac{\partial W_0}{\partial\tau}2\beta_1+\frac{\partial F(W_0,\ y,\ t)}{\partial y}\Big|_{y=-\psi}\frac{\tau-S_1}{\beta}$$
$$+W_1\frac{\partial F(W_0,\ -\psi,\ t)}{\partial W_0}\Big\}+W_0^{(\lambda+1)/\lambda}\Big\{\frac{\partial W_0}{\partial\tau}\frac{\partial S_1}{\partial t}$$
$$+u(-\psi,\ t)\frac{\partial W_0}{\partial\tau}\frac{\partial S_1}{\partial y}-\frac{\partial^2 W_0}{\partial\tau^2}D(-\psi,\ t)2\beta\frac{\partial S_1}{\partial y}\Big\}=0. \tag{3.5.4.9}$$

In this equation we have used the following expressions for the derivatives of $S(y, t)$:

$$\frac{\partial S}{\partial t}=\Big[\beta\frac{d\psi}{dt}+\left(\frac{d\beta}{dt}+2\beta_1\frac{d\psi}{dt}\right)\varepsilon\frac{\tau-S_1}{\beta}+\varepsilon\frac{\partial S_1}{\partial t}$$
$$+O((\varepsilon(\tau-S_1))^2)\Big]\Big|_{\tau=S/\varepsilon},$$
$$\frac{\partial S}{\partial y}=\Big[\beta+2\beta_1\varepsilon\frac{\tau-S_1}{\beta}+\varepsilon\frac{\partial S_1}{\partial x}+O((\varepsilon(\tau-S_1))^2)\Big]\Big|_{\tau=S/\varepsilon},$$
$$\left(\frac{\partial S}{\partial y}\right)^2=\Big[\beta^2+2\beta\left(\varepsilon^2\beta_1\frac{\tau-S_1}{\beta}+\varepsilon\frac{\partial S_1}{\partial x}\right)$$
$$+O((\varepsilon(\tau-S_1))^2\Big]\Big|_{\tau=S/\varepsilon},$$
$$\frac{\partial^2 S}{\partial y^2}=[2\beta_1+O(\varepsilon(\tau-S_1))]|_{\tau=S/\varepsilon},$$
$$y+\psi=[\varepsilon(\tau-S_1)/\beta+O((\varepsilon(\tau-S_1))^2)]|_{\tau=S/\varepsilon}.$$

Let us study Eq. (3.5.4.8) in greater detail. If we put

$$W_0(t,\ \tau)=1+[A(-\psi,\ t)-1]\chi(\tau) \tag{3.5.4.10}$$

and assume that

$$\frac{\beta(t)(\partial\psi/\partial t+u(-\psi,\ t))}{(A(-\psi,\ t)-1)d(-\psi,\ t)}=-b=\mathrm{const}>0,$$
$$\frac{D(-\psi,\ t)\beta^2(t)}{(A(-\psi,\ t)-1)d(-\psi,\ t)}=-1, \tag{3.5.4.11}$$

then for the function $\chi(\tau)$ we get a simple-wave equation of the KPP type (see Section 3.2.1):

$$b\frac{d\chi}{d\tau}-\frac{d^2\chi}{d\tau^2}-\chi(1-\chi)=0. \tag{3.5.4.12}$$

We know that a solution $\chi(\tau)$ satisfying the conditions

$$\chi(+\infty)=1,\ \chi(-\infty)=0 \tag{3.5.4.13}$$

exists if

$$b>2. \tag{3.5.4.14}$$

Equations (3.5.4.11) must be considered as comprising a system of equations for finding the unknown functions β and ψ.

The second term W_1 in the expansion of the solution to Eq. (3.5.4.5) and β_1 are constructed in the same manner as in Section 3.6.2 (see below).

The function χ has the following estimates:

$$\chi \sim \exp(l\tau) - C_0 \exp(2l\tau) \text{ as } \tau \to -\infty. \tag{3.5.4.15}$$

We rewrite Eq. (3.5.4.9) in the form

$$\frac{\partial^2 W_1}{\partial \tau^2} - b\frac{\partial W_1}{\partial \tau} - W_1(1-2\chi) = \frac{f}{A(-\psi, t)-1}, \tag{3.5.4.16}$$

where

$$\begin{aligned} f = -\Big\{ & \frac{b}{m\lambda}\beta\frac{d\psi}{dt}\frac{\partial W_0}{\partial \tau}W_0^{-(\lambda+1)/\lambda} + \frac{\partial W_0}{\partial \tau}\left(\beta^{-1}\frac{d\beta}{dt} + \frac{2\beta_1}{\beta}\frac{d\psi}{dt}\right)(\tau - S_1) \\ & + \frac{\partial W_0}{\partial t} + \frac{\partial W_0}{\partial \tau}u(-\psi, t)\,2\beta_1\frac{\tau - S_1}{\beta} + \frac{\partial W_0}{\partial \tau}(\tau - S_1)\left.\frac{\partial u}{\partial y}\right|_{y=-\psi} \\ & - D(-\psi, t)\frac{\partial^2 W_0}{\partial \tau^2}4\beta_1(\tau - S_1) - \left.\frac{\partial D}{\partial y}\right|_{y=-\psi}\beta(\tau - S_1)\frac{\partial^2 W_0}{\partial \tau^2} \\ & - D(-\psi, t)\frac{\partial W_0}{\partial \tau}2\beta_1 - \beta\frac{\partial W_0}{\partial \tau}\left.\frac{\partial D}{\partial y}\right|_{y=-\psi} \\ & - d(-\psi, t)\left.\frac{\partial A}{\partial y}\right|_{y=-\psi}\frac{\tau - S_1}{\beta}(W_0 - 1) \\ & - \left.\frac{\partial d(y, t)}{\partial y}\right|_{y=-\psi}\frac{\tau - S_1}{\beta}(A(-\psi, t) - W_0)(W_0 - 1) \\ & + \frac{\partial W_0}{\partial \tau}\frac{\partial S_1}{\partial t} + u(-\psi, t)\frac{\partial W_0}{\partial \tau}\frac{\partial S_1}{\partial y} - \frac{\partial^2 W_0}{\partial \tau^2}D(-\psi, t)\,2\beta\frac{\partial S_1}{\partial y}\Big\}. \end{aligned}$$

Note (see Section 3.2.1 and Eq. (3.5.4.10)) that the following estimates hold true:

$$\begin{aligned} &\frac{\partial W_0}{\partial t} = O(\exp(l\tau)) \text{ as } \tau \to -\infty, \\ &W_0 - 1 = O(\exp(l\tau)) \text{ as } \tau \to -\infty. \end{aligned} \tag{3.5.4.17}$$

The Wronskian of Eq. (3.5.4.16) has the form

$$V = \exp(b/\tau). \tag{3.5.4.18}$$

The general solution to Eq. (3.5.4.16) is specified by the formula

$$\begin{aligned} W_1 = & -\frac{\partial W_0/\partial \tau}{A(-\psi, t)-1}\int_{-\infty}^{\tau}\left(\int_{-\infty}^{\tau'}\frac{f\,\partial W_0/\partial \tau''\,d\tau''}{V}\right)\frac{V\,d\tau'}{(\partial W_0/\partial \tau')^2} \\ & + C_1\frac{\partial W_0}{\partial \tau} + C_2\frac{\partial W_0}{\partial \tau}\int_{-\infty}^{\tau}\frac{V\,d\tau'}{(\partial W_0/\partial \tau')^2}. \end{aligned} \tag{3.5.4.19}$$

The function f has the following estimate:

$$f = O(\tau \exp(l\tau)) \text{ as } \tau \to -\infty. \tag{3.4.5.20}$$

Hence, just as in Section 3.4 (see also Section 3.6), for arbitrary β_1 and S_1 the integrals in (3.5.4.19) are divergent.

If in the right-hand side of Eq. (3.5.4.16) we nullify the sum of terms with the asymptotic behavior exp $(l\tau)$ as $\tau \to -\infty$, we get the following equation:

$$\begin{aligned}
&\frac{b}{m\lambda}\beta\frac{d\psi}{dt}(A(-\psi, t)-1)l \\
&-\left(\frac{1}{\beta}\frac{d\beta}{dt}+\frac{2\beta_1}{\beta}\frac{d\psi}{dt}\right)S_1 l+(A(-\psi, t)-1) \\
&+\frac{\partial A(-\psi, t)}{\partial t}-u(-\psi, t)2\beta_1(A(-\psi, t)-1)S_1 l/\beta \\
&-\frac{\partial u}{\partial y}\Big|_{y=-\psi} lS_1(A(-\psi, t)-1) \\
&+D(-\psi, t)l^2 4\beta_1 S_1(A(-\psi, t)-1)+\frac{\partial D}{\partial y}\Big|_{y=-\psi}\beta S_1 l^2(A(-\psi, t)-1) \\
&-\beta(A(-\psi, t)-1)\frac{\partial D}{\partial y}\Big|_{y=-\psi} l-D(-\psi, t)(A(-\psi, t)-1)2\beta l \\
&+d(-\psi, t)\frac{\partial A}{\partial y}\Big|_{y=-\psi} S_1(A(-\psi, t)-1)/\beta \\
&+\frac{\partial d(y, t)}{\partial y}\Big|_{y=-\psi}(A(-\psi, t)-1)(A(-\psi, t)-1)/\beta \\
&+\frac{\partial S_1}{\partial t}l(A(-\psi, t)-1)+u(-\psi, t)\frac{\partial S_1}{\partial y}(A(-\psi, t)-1)l \\
&-l^2 D(-\psi, t)2\beta\frac{\partial S_1}{\partial y}(A(-\psi, t)-1)=0.
\end{aligned} \tag{3.5.4.21}$$

If we nullify in (3.5.4.16) the sum of terms with the asymptotic behavior $\tau(\exp(l\tau))$ as $\tau \to -\infty$, we arrive at an equation for β_1:

$$\begin{aligned}
&\beta^{-1}\left(\frac{d\beta}{dt}+2\beta_1\frac{d\psi}{dt}\right)l+2\beta_1 lu(-\psi, t)/\beta+l\frac{\partial u}{\partial y}\Big|_{y=-\psi} \\
&-D(-\psi, t)4\beta_1 l^2-\frac{\partial D}{\partial y}\Big|_{y=-\psi}\beta l^2-\frac{\partial d(y, t)}{\partial y}\Big|_{y=-\psi}(A(-\psi, t)-1)/\beta \\
&-d(-\psi, t)\frac{\partial A}{\partial y}\Big|_{y=-\psi}/\beta=0.
\end{aligned} \tag{3.5.4.22}$$

Combining (3.5.2.21) and (3.5.2.22) results in the following equation for S_1:

$$\begin{aligned}
&\frac{\partial A(-\psi, t)}{\partial t}+\frac{b}{m\lambda}\beta\frac{d\psi}{dt}(A(-\psi, t)-1)l-\beta(A(-\psi, t)-1)l\frac{\partial D}{\partial y}\Big|_{y=-\psi} \\
&-D(-\psi, t)2\beta l(A(-\psi, t)-1)+\frac{\partial S_1}{\partial t}l(A(-\psi, t)-1) \\
&+u(-\psi, t)\frac{\partial S_1}{\partial y}l(A(-\psi, t)-1)-l^2 D(-\psi, t)2\beta\frac{\partial S_1}{\partial y}=0.
\end{aligned} \tag{3.5.4.23}$$

Solving (3.5.4.22), we can find β_1 (by employing (3.5.4.11)).

$$\beta_1=\left[\frac{l}{\beta}\frac{\partial\beta}{dt}+l\left.\frac{\partial u}{\partial y}\right|_{y=-\psi}-\left.\frac{\partial D}{\partial y}\right|_{y=-\psi}\beta l^2-\left.\frac{\partial d(y,\ t)}{\partial y}\right|_{y=-\psi}\frac{S_1(A(-\psi,t)-1)}{\beta}\right.$$
$$\left.-d(-\psi,\ t)\left.\frac{\partial A}{\partial y}\right|_{y=-\psi}/\beta\right][D(-\psi,\ t)2l\sqrt{b^2-4}]^{-1}. \qquad (3.5.4.24)$$

This implies that $b>2$, just as in Section 3.2.1.

Allowing for Eqs. (3.5.4.21) and (3.5.4.22) yields a new estimate for function f as $\tau\to-\infty$:

$$f=O(\tau\exp(2l\tau)),\ fV^{-1}\frac{\partial W_0}{\partial\tau}=O(\tau\exp\{(3l-b)\tau\}),$$
$$V\Big/\left(\frac{dW_0}{d}\right)^2=O(\exp\{(b-2l)\tau\}).$$

Let us analyze the exponent in the last estimate. Equation (3.5.4.12) implies that $l=(b-\sqrt{b^2-4})/2$, so that $b-2l=\sqrt{b^2-4}>0$. Hence the integrals in (3.5.4.19) exist and, by virtue of the boundary condition for W_1 at $\tau=0$,

$$C_2=\frac{1}{A(-\psi,\ t)-1}\int_{-\infty}^{0}V^{-1}f\frac{\partial W_0}{\partial\tau}\,d\tau,\quad C_1=0.$$

By virtue of the boundary condition at $y=-\psi(t)$,

$$W_0(y,\ t)|_{y=-\psi(t)}=1+[A(-\psi,\ t)-1]\,\chi(S_1(-\psi,\ t))=v_0(t).$$

The last equation can be used to determine $S_1(-\psi,\ t)$, the boundary condition (3.5.4.7) for Eq. (3.5.4.23). After carrying out the necessary substitutions and transformations we arrive at the final expression for the function W_1 specified in the hypothesis of Theorem 3.5.4.1, with $W_1=O(\tau\exp(2l\tau))$ as $\tau\to-\infty$. The proof of the theorem is complete.

Example 3.5.4.1 If d, u, A, and D are constant, that is, if the medium has not been excited, then

$$v_0(t)=(b/a)^\lambda,\quad \chi(0)=[(b/a)^\lambda-1](A-1)^{-1},$$
$$\tau=y+t,\quad \psi=b_1t,\quad \beta=1,\quad b=(b_1+u)/[(A-1)\,d].$$

The function φ can be expressed in terms of v (see (3.5.4.2)).

Thus, the rate at which the interphase is formed depends on the rate of admission u of the substance, the reaction rate, and the diffusion coefficient. Equation (3.5.4.11) yields

$$b_1=(1-A)\,bd-u,\quad D=(1-A)\,d.$$

The graphs of these solutions v and φ are similar to those depicted in Figures 3.5. and 3.6.

3.6 Diffusion of Light in an Active Medium

In this section we study the asymptotic solutions for a semi-linear hyperbolic equation and compare them with similar solutions for semi-linear parabolic equations. For a discussed classe of perturbations in th medium there can be no asymptotic solution to a semi-linear parabolic equation (the KPP problem) at the minimal speed of propagation of the wave.

3.6.1 Diffusion of Light in an Active Medium

The hyperbolic diffusion equation (the heat equation or telegrapher's equation) emerges in the following problems: (a) diffusion of light in an active medium, (b) propagation of thermal waves in continuous media, (c) turbulent diffusion, (d) the random walk of particles described by the Fokker-Planck equation (K. H. Cramer and S. Chandrasekhar), and in other problems (for a complete bibliography on the subject see [3.3, 3.34]).

Generally speaking, a hyperbolic transport equation appears in problems in which we cannot assume the diffusion rate to be infinitely high and the mean free path of particles infinitely small and/or in which there are high gradients of temperature, concentration, and intensity (of the quantity transferred).

In dimensionless form, for large values of time $t = \tilde{t}\,\varepsilon$, the problem is formulated as follows:

$$Lu \stackrel{\text{def}}{=} \varepsilon \frac{\partial u}{\partial t} + \varepsilon^2 \frac{\partial^2 u}{\partial t^2} - \varepsilon^2 \frac{\partial}{\partial x}\left(\lambda(x, t) K(u) \frac{\partial u}{\partial x}\right) - F(u) = 0,$$

$$x \in R,\ t \in [0, T],\ 0 \leqslant \varepsilon < 1,\ u > 0,\ K(u) > 0, \quad (3.6.1.1)$$

$$F(a_i) = 0,\ a_i = \text{const},\ i = 0,\ 1,\ a_1 > a_0,\ dF/du|_{a_i} \neq 0.$$

Here ε is a small parameter, which appears naturally if we consider the solution to Eq. (3.6.1.1) for large $t = \tilde{t}/\varepsilon$ (what interests specialists in all applications is whether the solution of a perturbation problem "reaches" the limiting wave, which propagates with the minimal speed; see [3.25-3.29]); F is the source distribution function; the functions $F(u)$, $K(u)$, $K(u)\,\partial u/\partial x$, and $\lambda(x, t)$ are infinitely differentiable and positive (the first three for $u \leqslant a_1$); and the function $\lambda(x, t) \geqslant \delta = \text{const} > 0$ characterizes the slowly varying properties of the medium. The boundary conditions are

$$u|_{x \to +\infty} \to a_1, \quad u|_{x \to -\infty} \to a_0. \quad (3.6.1.2)$$

The solution to the problem depends essentially on whether equation $K(u) = b$ has a root.

Remark In what follows we consider all the cases. Variant (3.6.1.3) can be reduced to the KPP equation (see p. 333) and corresponds to the case when the standard equation has an asymptotic solution of the (3.2.1.11) type. In some instances [3.43] when there is an asymptotic solution with the property $\Theta \sim \tau \exp(b_{\min}\tau/2)$, $b = b_{\min}$, $\tau \to -\infty$, the method given in Sec. 3.6.3 can be used.

The following variants are possible:

(1) $K(u) \neq b^2, \quad u \in [a_0, a_1],$ (3.6.1.3)

(2) $K(a_0) = b^2,$ or (3.6.1.4)

(3) $K(a_1) = b^2,$ (3.6.1.5)

where b is the speed of propagation of the wave.

Moreover, the manner in which $K(u)$ depends on u also has a strong effect on the properties of the solution.

For one dimension, variants (3.6.1.4) and (3.6.1.5) have been extensively treated in [3.3], while for many dimensions the solution for these variants is constructed in Sections 3.6.4 and 3.6.5. In the present section we study the case of (3.6.1.3), where all the assertions remain valid for $K = 1$, too.

Definition 3.6.1.1 A nonnegative continuous solution $u(x, t, \varepsilon)$ to Eq. (3.6.1.1) is said to be a formed wave if

$$a_0 \leqslant u(x, t, \varepsilon) \leqslant a_1; \quad u(x, t, \varepsilon) = a_0 + \delta_1(\varepsilon), \ x \in \Omega_1;$$
$$u(x, t, \varepsilon) = a_1 - \delta_2(\varepsilon), \quad x \in \Omega_2; \quad \delta_i(\varepsilon) = O(\varepsilon^m), \ m > 0, \ i = 1, 2,$$

where Ω_i are regions in R.

We will seek the asymptotic solution to problem (3.6.1.1), (3.6.1.2) in the form

$$u(x, t, \varepsilon) = \chi(\tau) + \varepsilon W_1(t, \tau) + O(\varepsilon^2)|_{\tau = S(x, t, \varepsilon)/\varepsilon}. \quad (3.6.1.6)$$

Theorem 3.6.1.1 *Problem* (3.6.1.1), (3.6.1.2) *has an asymptotic solution of the* (3.6.1.6) *type, with function* $\chi(\tau)$ *satisfying the simple-wave equation*

$$b\frac{d\chi}{d\tau} - \frac{d}{d\tau}\left((K(\chi) - b^2)\frac{d\chi}{d\tau}\right) - F(\chi) = 0, \quad (3.6.1.7)$$
$$\chi|_{\tau\to-\infty} = a_0 + 0, \ \chi|_{\tau\to+\infty} = a_1 - 0, \ K(\chi) - b^2 > 0.$$

The following inequalities hold true:

$$dR/d\chi|_{\chi=a_0} > 0, \quad dR/d\chi|_{\chi=a_1} < 0,$$

with $R = F(\chi)(K(\chi) - b^2) > 0, \quad \chi \in (a_0, a_1),$ *and* $b > b_{\min} = 2\{[(dF/d\chi|_{\chi=a_0})K(a_0)] \times [1 + 4(dF/d\chi|_{\chi=a_0})]^{-1}\}^{1/2}$, *and the following estimates hold true* (*see the Remark on p. 320*):

$$\chi(\tau) \sim a_0 + \exp\frac{l\tau}{K(a_0) - b^2} \ \ as \ \ \tau \to -\infty,$$
$$l = \left(b - \sqrt{(b^2 - b^2_{\min})(1 + 4(dF/d\chi|_{\chi=a_0}))}\right)/2;$$
$$\chi(\tau) \sim a_1 - \exp\frac{-l_1\tau}{K(a_1) - b^2} \ \ as \ \ \tau \to +\infty,$$
$$l_1 = -b/2 + \sqrt{b^2/4 + dF/d\chi|_{\chi=a_1}(K(a_1) - b^2)} > 0.$$

21—0105

The function $S\ (x,\ t,\ \varepsilon)$ *has the form*

$$S(x,\ t,\ \varepsilon)=\beta(t)(x+\varphi(t))+\beta_1(x+\varphi(t))^2+\varepsilon S_1(x,\ t),$$

where the functions β *and* φ *can be found by solving the system of equations*

$$\beta(t)=(\lambda(-\varphi,\ t))^{-1/2},\quad \beta d\varphi/dt=b=\mathrm{const}>0. \qquad (3.6.1.8)$$

The function $\beta_1\ (t)$ *is specified by the following formula:*

$$\begin{aligned}\beta_1(t)=\Big\{&-\beta^{-1}\frac{d\beta}{dt}\\ &-2l\,\frac{d\varphi}{dt}\,\frac{d\beta}{dt}\,\frac{1}{K(a_0)-b^2}+\beta\,\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}\frac{lK(a_0)}{K(a_0)-b^2}\Big\}\\ &\times\{-2\lambda(-\varphi,\ t)\,[b+2l]\}^{-1}.\end{aligned} \qquad (3.6.1.9)$$

The function S_1 *can be found by solving the equation*

$$-\frac{\partial S_1}{\partial t}-\frac{2l\beta}{K(a_0)-b^2}\,\frac{d\varphi}{dt}\,\frac{\partial S_1}{\partial t}-\Big[\frac{d}{dt}\Big(\beta\frac{d\varphi}{dt}\Big)+2\beta_1\Big(\frac{d\varphi}{dt}\Big)^2\Big]$$
$$+\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}K(a_0)+2\lambda(-\varphi,\ t)\,\frac{lK(a_0)}{K(a_0)-b^2}\,\beta\,\frac{\partial S_1}{\partial x}=0.$$

The function W_1 *has the form*

$$W_1=C_1\frac{d\chi}{d\tau}+\frac{1}{\lambda(-\varphi,\ t)\,\beta^2}\,\frac{d\chi}{d\tau}\int\limits_{-\infty}^{\tau}V\Big(\frac{d\chi}{d\tau'}\Big)^{-2}\Big(\int\limits_{+\infty}^{\tau'}\frac{f\,(d\chi/d\tau)\,d\tau}{V(K(\chi)-b^2)}\Big)\,d\tau'.$$

The functions f *and* V *will be defined later (see* (3.6.1.11′) *and* (3.6.1.13′)) *and the following estimates hold true:*

$$W_1=O\Big(\exp\frac{(b-l)\,\tau}{K(a_0)-b^2}\Big)\quad as\ \ \tau\to-\infty,$$

$$b_{\min}<b<\frac{3b_{\min}\sqrt{1+4\,dF/d\chi\,|_{\chi=a_0}}}{\sqrt{8+36\,dF/d\chi\,|_{\chi=a_0}}}\ ;$$

$$W_1=O\Big(\tau^2\exp\frac{(b-l)\,\tau}{K(a_0)-b^2}\Big)\quad as\ \ \tau\to-\infty,$$

$$b=\frac{3\sqrt{1+4\,(dF/d\chi\,|_{\chi=a_0})}}{\sqrt{8+36\,(dF/d\chi\,|_{\chi=a_0})}}\ ;$$

$$W_1=O\Big(\tau\exp\frac{2l\tau}{K(a_0)-b^2}\Big)\quad as\ \ \tau\to-\infty,$$

$$b>\frac{3\sqrt{1+4\,(dF/d\chi\,|_{\chi=a_0})}}{\sqrt{8+36\,(dF/d\chi_{\chi=a_0})}}\ ;$$

and

$$W_1=O\Big(\tau^2\exp\frac{-l\tau}{K(a_1)-b^2}\Big)\quad as\ \ \tau\to+\infty.$$

Proof. The main equations obtained after we substitute solutions (3.6.1.6) into Eq. (3.6.1.1) have the form

$$\beta \frac{d\varphi}{dt} \frac{d\chi}{d\tau} + \left(\beta \frac{d\varphi}{dt}\right)^2 \frac{d^2\chi}{d\tau^2} - \beta^2\lambda(-\varphi, t) \frac{d}{d\tau} \left(K(\chi) \frac{d\chi}{d\tau}\right) - F(\chi) = 0, \tag{3.6.1.10}$$

$$\beta \frac{d\varphi}{dt} \frac{dW_1}{d\tau} + \left(\beta \frac{d\varphi}{dt}\right)^2 \frac{d^2W_1}{d\tau^2} - \beta^2\lambda(-\varphi, t) \frac{d^2}{d\tau^2} (K(\chi) W_1) - \frac{dF(\chi)}{d\chi} W_1 = f, \tag{3.6.1.11}$$

where

$$\begin{aligned} f = & -\frac{d\chi}{d\tau} \left[\frac{\tau - S_1}{\beta} \left(\frac{d\beta}{dt} + 2\beta_1 \frac{d\varphi}{dt}\right) + \frac{\partial S_1}{\partial t}\right] \\ & - \frac{\partial^2\chi}{\partial\tau^2} \left[2\beta \frac{d\varphi}{dt} \frac{\partial S_1}{\partial t} + 2(\tau - S_1) \frac{d\varphi}{dt} \left(\frac{\partial\beta}{dt} + 2\beta_1 \frac{d\varphi}{\partial t}\right)\right] \\ & - \frac{d\chi}{d\tau} \left[\frac{d}{dt} \left(\beta \frac{d\varphi}{dt}\right) + 2\beta_1 \left(\frac{d\varphi}{dt}\right)^2\right] \\ & + \beta \frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi} (\tau - S_1) \frac{\partial}{\partial\tau} \left(K(\chi) \frac{d\chi}{d\tau}\right) + \beta \frac{\partial\chi}{\partial x}\Big|_{x=-\varphi} K(\chi) \frac{d\chi}{d\tau} \\ & + 2\beta_1\lambda(-\varphi, t) K(\chi) \frac{d\chi}{d\tau} + 4\beta_1\lambda(-\varphi, t)(\tau - S_1) \frac{\partial}{\partial\tau} \left(K(\chi) \frac{d\chi}{dt}\right) \\ & + \lambda(-\varphi, t) \frac{\partial}{\partial\tau} \left(K(\chi) \frac{d\chi}{d\tau}\right) \left(2\beta \frac{\partial S_1}{\partial x}\right). \end{aligned} \tag{3.6.1.11'}$$

Here we have used the following relationships:

$$\begin{aligned} \frac{\partial S}{\partial t} &= \beta \frac{d\varphi}{dt} + \varepsilon \left(\frac{d\beta}{dt} + 2\beta_1 \frac{d\varphi}{dt}\right) \frac{\tau - S_1}{\beta} + \varepsilon \frac{\partial S_1}{\partial t} + O((\tau\varepsilon)^2)\Big|_{\tau=S/\varepsilon}, \\ \frac{\partial^2 S}{\partial t^2} &= \frac{d}{dt} \left(\beta \frac{d\varphi}{dt}\right) + 2\beta_1 \left(\frac{d\varphi}{dt}\right)^2 + O(\varepsilon\tau)\Big|_{\tau=S/\varepsilon}, \\ \frac{\partial S}{\partial x} &= \beta + 2\frac{\beta_1}{\beta} \varepsilon(\tau - S_1) - \varepsilon \frac{\partial S_1}{\partial x} + O((\varepsilon\tau)^2)\Big|_{\tau=S/\varepsilon}, \\ \frac{\partial^2 S}{\partial x^2} &= 2\beta_1 + O((\varepsilon\tau))\Big|_{\tau=S/\varepsilon}, \\ x + \varphi &= \frac{\varepsilon(\tau - S_1)}{\beta} + O((\varepsilon\tau)^2)\Big|_{\tau=S/\varepsilon}. \end{aligned}$$

The function χ solves the equation

$$b \frac{d\chi}{dt} - \frac{d}{d\tau} \left[(K(\chi) - b^2) \frac{d\chi}{d\tau}\right] - F(\chi) = 0, \tag{3.6.1.12}$$

and the functions $\beta(t)$ and $\varphi(t)$ satisfy system (3.6.1.8). Hence, for $Lu = O(\varepsilon^2)$ to be valid it is sufficient that (3.6.1.11) be valid and that

$$W_1|_{\tau\to-\infty} = 0, \quad W_1|_{\tau\to+\infty} = 0.$$

The general solution to Eq. (3.6.1.11) has the form

$$W_1 = C_1\frac{d\chi}{d\tau} + C_2\frac{d\chi}{d\tau}\int_{-\infty}^{\tau} V\left(\frac{d\chi}{d\tau}\right)^{-2} d\tau$$

$$+\frac{1}{\beta^2\lambda(-\varphi, t)}\frac{d\chi}{d\tau}\int_{-\infty}^{\tau} V\left(\frac{d\chi}{d\tau'}\right)^{-2}\left(\int_{+\infty}^{\tau'}\frac{f(d\chi/d\tau)\,d\tau}{V(K(\chi)-b^2)}\right)d\tau'. \qquad (3.6.1.13)$$

The function V, or the Wronskian of Eq. (3.6.1.11), has the form

$$V = (K(\chi)-b^2)^{-2}\exp\left(b\int_{-\infty}^{\tau}\frac{d\tau}{K(\chi)-b^2}\right). \qquad (3.6.1.13')$$

Since, by the hypothesis of Theorem 3.6.1.1, $K(a_0) - b^2$ is positive and in view of the results established in Section 3.2, the function χ has the following asymptotic behavior as $\tau\to-\infty$:

$$\chi \sim a_0 + \exp\bar{l}\tau + \exp 2\bar{l}\tau + \ldots, \quad \bar{l} = \frac{l}{K(a_0)-b^2},$$

where

$$l = b/2 - \sqrt{b^2/4 - (K(a_0)-b^2)(dF/d\chi|_{\chi=a_0})}$$
$$= \frac{1}{2}\left(b - \mu\sqrt{b^2 - b_{\min}^2}\right) > 0,$$
$$\mu = \sqrt{1 + 4(dF/d\chi|_{\chi=a_0})}$$

(see Section 3.2).

The minimal speed can be found from the condition that the discriminant be nonnegative:

$$b < b_{\min} = 2\sqrt{\frac{K(a_0)(dF/d\chi|_{\chi=a_0})}{1+4(dF/d\chi|_{\chi=a_0})}}.$$

The following estimates hold true:

$$f = O(\tau\exp(\bar{l}\tau)) \quad \text{as} \quad \tau\to-\infty,$$
$$V = O\left(\exp\frac{b\tau}{K(a_0)-b^2}\right) \quad \text{as} \quad \tau\to-\infty,$$

and the integrals in (3.6.1.13) are divergent for arbitrary β_1 and S_1.

Nullifying the sum of coefficients of $\exp(\bar{l}\tau)$ as $\tau \to -\infty$, we arrive at the following equation

$$-\bar{l}\left[\frac{\partial S_1}{\partial t} - \frac{S_1}{\beta}\left(\frac{d\beta}{dt} + 2\beta_1\frac{d\varphi}{dt}\right)\right]$$
$$-\bar{l}^2\left[2\beta\frac{d\varphi}{dt}\frac{\partial S_1}{\partial t} - 2S_1\frac{d\varphi}{dt}\left(\frac{d\beta}{dt} + 2\beta_1\frac{d\varphi}{dt}\right)\right]$$
$$-\bar{l}\left[\frac{d}{dt}\left(\beta\frac{d\varphi}{dt}\right) + 2\beta_1\left(\frac{d\varphi}{dt}\right)^2\right] - \beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi} K(a_0)\bar{l}^2$$
$$+\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi} K(a_0)\bar{l} + 2\beta_1\lambda(-\varphi, t)K(a_0)\bar{l} - 4\beta_1\lambda(-\varphi, t)S_1K(a_0)\bar{l}^2$$
$$+2\lambda(-\varphi, t)K(a_0)\bar{l}^2\beta\frac{\partial S_1}{\partial x} = 0. \tag{3.6.1.14}$$

Nullifying the sum of coefficients of $\tau\exp(\bar{l}\tau)$ as $\tau \to -\infty$, we get

$$\begin{gathered}-\frac{\bar{l}}{\beta}\left(\frac{d\beta}{dt} + 2\beta_1\frac{d\varphi}{dt}\right) - 2\bar{l}^2\frac{d\varphi}{dt}\left(\frac{d\beta}{dt} + 2\beta_1\frac{d\varphi}{dt}\right)\\ +\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}\bar{l}^2K(a_0) + 4\beta_1\lambda(-\varphi, t)\bar{l}^2K(a_0) = 0,\end{gathered} \tag{3.6.1.15}$$

from which Eq. (3.6.1.9) follows.

Combining (3.6.1.14) with (3.6.1.15) yields an equation for finding S_1:

$$\frac{\partial S_1}{\partial t} - \bar{l}2\beta\frac{d\varphi}{dt}\frac{\partial S_1}{\partial t} - \left[\frac{d}{dt}\left(\beta\frac{d\varphi}{dt}\right) + 2\beta_1\left(\frac{d\varphi}{dt}\right)^2\right]$$
$$+\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}K(a_0) + 2\lambda(-\varphi, t)\bar{l}K(a_0)\beta\frac{\partial S_1}{\partial x} = 0. \tag{3.6.1.15'}$$

Let us investigate the denominator in (3.6.1.9) (see also (3.6.1.15)). Allowing for Eq. (3.6.1.8), we get

$$-\frac{2}{\beta}\frac{d\varphi}{dt} - 4\bar{l}\left(\frac{d\varphi}{dt}\right)^2 + 4\lambda(-\varphi, t)\bar{l}K(a_0) \tag{3.6.1.16}$$
$$= 2\lambda(-\varphi, t)[-b - 2\bar{l}b^2 + 2\bar{l}K(a_0)] = -2\lambda(-\varphi, t)[b - 2\bar{l}] \leqslant 0.$$

We see that the denominator can vanish.

Let us continue investigating (3.6.1.13). Equation (3.6.1.11) implies that if conditions (3.6.1.14) and (3.6.1.15) are met, the function f has an estimate

$$f = O(\tau\exp 2\bar{l}\tau) \text{ as } \tau \to -\infty. \tag{3.6.1.17}$$

Moreover, as $\tau \to -\infty$, the following estimates hold true:

$$f\frac{d\chi}{d\tau}V^{-1} = O\left(\tau\exp\left\{\left(3\bar{l} - \frac{b}{K(a_0) - b^2}\right)\tau\right\}\right),$$
$$V\left(\frac{d\chi}{d\tau}\right)^{-2} = O\left(\exp\left\{\left(\frac{b}{K(a_0) - b^2} - 2\bar{l}\right)\tau\right\}\right).$$

Let us examine the behavior of the exponent in the last expression as $\tau \to -\infty$:

$$\frac{b}{K(a_0)-b^2} - 2\bar{l} = \frac{b-2l}{K(a_0-b^2)}$$
$$= \frac{2\sqrt{b^2/4-(K(a_0)-b^2)(dF/d\chi|_{\chi=a_0})}}{K(a_0)-b^2} > 0, \qquad (3.6.1.18)$$

where we have employed an equation that follows from (3.6.1.12) and from the results obtained in Section 3.2.1, namely

$$b = l + \frac{K(a_0)-b^2}{l} \left.\frac{dF}{d\chi}\right|_{\chi=a_0},$$

with

$$l = \lim_{\tau\to-\infty} \left(\frac{d^2\Theta}{d\xi^2} \Big/ \frac{d\Theta}{d\xi}\right) > 0.$$

Condition (3.6.1.18) implies that

$$3\bar{l} - \frac{b}{K(a_0)-b^2} = \left(\frac{b}{2} - \frac{3}{2}\mu\sqrt{b^2-b_{\min}^2}\right) \Big/ (K(a_0)-b^2) > 0,$$

with

$$b_{\min} < b < \frac{3b_{\min}(1+4(dF/d\chi|_{\chi=a_0}))}{\sqrt{8+36(dF/d\chi|_{\chi=a_0})}}.$$

Hence, the inner integral in (3.6.1.13) has the estimate

$$I = \int^{\tau} \frac{f(d\chi/d\tau)\,d\tau}{\sqrt{(K(\chi)-b^2)}} \sim O\left(\tau \exp\frac{(3l-b)\tau}{K(a_0)-b^2}\right) \qquad (3.6.1.19)$$

and, as $\tau \to -\infty$, is convergent for

$$b_{\min} < b < \frac{3b_{\min}\sqrt{1+4(dF/d\chi|_{\chi=a_0})}}{\sqrt{8+36(dF/d\chi|_{\chi=a_0})}}$$

and divergent for $b \geqslant (3/\sqrt{8})\,b_{\min}$.

The following estimate holds true:

$$I = O(\tau^2),\ b = \frac{3b_{\min}\sqrt{1+4(dF/d\chi|_{\chi=a_0})}}{\sqrt{8+36(dF/d\chi|_{\chi=a_0})}}, \qquad (3.6.1.20)$$

and, as $\tau \to -\infty$, the following estimates hold true:

$$I_0 = \int_{-\infty}^{\tau} \sqrt{\left(\frac{d\chi}{d\tau}\right)^{-2}} \left(\int_{+\infty}^{\tau'} \frac{f(d\chi/d\xi)\,d\xi}{\sqrt{(K(\chi)-b^2)}}\right) d\tau'$$
$$= O\left(\exp\frac{(b-2l)\tau}{K(a_0)-b^2}\right),\ b-2l > 0,$$

for

$$b_{\min} < b < \frac{3b_{\min}\mu}{\sqrt{8+36\,(dF/d\chi\,|_{\chi=a_0})}}\,;$$

$$I_0 = O\left(\tau^2 \exp\frac{(b-2l)\,\tau}{K(a_0)-b^2}\right) \text{ at } b = \frac{3b_{\min}\mu}{\sqrt{8+36\,(dF/d\chi\,|_{\chi=a_0})}}\,;$$

$$I_0 = O\left(\tau \exp\frac{l\tau}{K(a_0)-b^2}\right) \text{ for } b > \frac{3b_{\min}\mu}{\sqrt{8+36\,(dF/d\chi\,|_{\chi=a_0})}}\,.$$

Thus, $C_2 \equiv 0$ in (3.6.1.13), which implies the validity of (3.6.1.10) and the following estimates (as $\tau \to -\infty$):

$$W_1 = O\left(\exp\frac{(b-l)\,\tau}{K(a_0)-b^2}\right),\ b_{\min} < b < \frac{3\mu b_{\min}}{\sqrt{8+36\,(dF/d\chi\,|_{\chi=a_0})}}\,;$$

$$W_1 = O\left(\tau^2 \exp\frac{(b-l)\,\tau}{K(a_0)-b^2}\right),\ b = \frac{3\mu b_{\min}}{\sqrt{8+36\,(dF/d\chi\,|_{\chi=a_0})}}\,; \tag{3.6.1.21}$$

$$W_1 = O\left(\tau \exp\frac{2l\tau}{K(a_0)-b^2}\right),\quad b > \frac{3\mu b_{\min}}{\sqrt{8+36\,(dF/d\chi\,|_{\chi=a_0})}}\,.$$

Let us study the asymptotic behavior of W_1 as $\tau \to +\infty$. The estimate

$$\chi(\tau) = a_1 - \exp(-\bar{l}_1\tau) + o(\exp(-\bar{l}_1\tau)) \text{ as } \tau \to +\infty,$$

$$\bar{l}_1 = \frac{l_1}{K(a_1)-b^2}\,,\quad l_1 = -b/2 + \sqrt{b^2/4 + (dF/d\chi|_{\chi=a_1})\,(K(a_1)-b^2)} > 0,$$

implies that

$$I_1 = \int_{+\infty}^{\tau} \frac{f\,(d\chi/d\tau)\,d\tau}{V\,(K(\chi)-b^2)} \sim O\left(\tau \exp\frac{-(2l_1+b)\,\tau}{K(a_1)-b^2}\right) \text{ as } \tau \to +\infty,$$

$$V\left(\frac{d\chi}{d\tau}\right)^{-2} = O\left(\exp\frac{(b+2l_1)\,\tau}{K(a_1)-b^2}\right) \text{ as } \tau \to +\infty,$$

and the integral I_1 converges as $\tau \to +\infty$. Then we have the following estimate:

$$W_1 = O\left(\tau^2 \exp\frac{-l\tau}{K(a_1)-b^2}\right) \text{ as } \tau \to +\infty. \tag{3.6.1.22}$$

We have, therefore, constructed an asymptotic solution to problem (3.6.1.1), (3.6.1.2), so that the proof of Theorem 3.6.1.1 is complete.

The solution to problem (3.6.1.1), (3.6.1.2) is stable with respect to smooth perturbations and can be compared to the solution to a similar problem for the parabolic equation in Section 3.6.3 (see below).

3.6.2 Diffusion of Light in an Active Medium (Continued)

In this section we consider another variant of the model used to describe the propagation of light in an active medium.

Let us again take up problem (3.6.1.1), (3.6.1.2). Suppose that the algebraic equation (3.6.1.3) has no roots $K(u) \neq b^2$ for $u \in (a_0, a_1)$ and that $K(a_0) - b^2 < 0$, for the sake of definiteness. The asymptotic solution to problem (3.6.1.1), (3.6.1.2) is of the (3.6.1.6) type. The following theorem is valid:

Theorem 3.6.2.1 *Problem* (3.6.1.1), (3.6.1.2) *has an asymptotic solution of the* (3.6.1.6) *type, with* $\chi(\tau)$ *satisfying the simple-wave equation*

$$b\frac{d\chi}{d\tau}+\frac{d}{d\tau}\left(|K(\chi)-b^2|\frac{d\chi}{d\tau}\right)-F(\chi)=0,$$
$$\chi|_{\tau\to+\infty}=a_0,\quad \chi|_{\tau\to-\infty}=a_1,\quad K(\chi)-b^2<0. \tag{3.6.2.1}$$

The following inequalities hold true:

$$R=F(\chi)(K(\chi)-b^2)<0,\ \frac{dR}{d\chi}\Big|_{\chi=a_0}<0,\ \frac{dR}{d\chi}\Big|_{\chi=a_1}>0$$

for $\chi\in(a_0, a_1)$ *and*

$$b<b_{\max}=-2\sqrt{\frac{|dF/d\chi|_{\chi=a_1}|K(a_1)|}{4|dF/d\chi|_{\chi=a_1}|-1}}, \tag{3.6.2.2}$$

and the following estimates hold true (see the Remark on p. 320):

$$\chi\sim a_1-\exp\frac{l_1\tau}{|K(a_1)-b^2|}\quad as\ \tau\to-\infty,$$
$$l_1=-b/2+\sqrt{b^2/4-|dF/d\chi|_{\chi=a_1}(K(a_1)-b^2)|}>0;$$
$$\chi\sim a_0+\exp\frac{-l_0\tau}{|K(a_0)-b^2|}\quad as\ \tau\to+\infty,$$
$$l_0=b/2+\sqrt{b^2/4+(dF/d\chi|_{\chi=a_0})|K(a_0)-b^2|}>0.$$

The function $S(x, t, \varepsilon)$ *has the form*

$$S(x, t, \varepsilon)=\beta(t)(x+\varphi(t))+\beta_1(t)(x+\varphi(t))^2+\varepsilon S_1(x, t) \tag{3.6.2.3}$$

where the functions β *and* φ *can be found from the system of equations*

$$\beta(t)=\frac{1}{\sqrt{\lambda(-\varphi, t)}},\quad \beta\frac{d\varphi}{dt}=b=\text{const}<0, \tag{3.6.2.3'}$$

the function β_1 *is defined as follows*:

$$\beta_1(t)=\left\{\beta^{-1}\frac{d\beta}{dt}-2l_0\frac{d\varphi}{dt}\frac{d\beta}{dt}\frac{1}{|K(a_0)-b^2|}\right.$$
$$\left.+\beta\frac{\partial\lambda}{\partial x}\Big|_{x=\varphi}\frac{l_0K(a_0)}{|K(a_0)-b^2|}\right\}[2\lambda(-\varphi, t)(b-2l_0)]^{-1}, \tag{3.6.2.3''}$$

and the function S_1 can be determined by solving the equation

$$\frac{\partial S_1}{\partial t}-\frac{2l_0\beta}{|K(a_0)-b^2|}\frac{d\varphi}{dt}\frac{\partial S_1}{\partial t}+\left[\frac{d}{dt}\left(\beta\frac{d\varphi}{dt}\right)+2\beta_1\left(\frac{d\varphi}{dt}\right)^2\right]$$
$$-\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}K(a_0)+2\lambda(-\varphi,t)\frac{l_0K(a_0)}{|K(a_0)-b^2|}\beta\frac{\partial S_1}{\partial x}=0.$$

The function W_1 has the form

$$W_1=C_1\frac{d\chi}{d\tau}-\frac{1}{\lambda(-\varphi,t)\beta^2}\frac{d\chi}{d\tau}\int\limits_{+\infty}^{\tau}V\left(\frac{d\chi}{d\eta}\right)^{-2}\left(\int\limits_{+\infty}^{\eta}\frac{f(d\chi/d\tau)\,d\tau}{V(K(\chi)-b^2)}\right)d\eta,$$

where f and V are defined in (3.6.1.11′) *and* (3.6.1.6) (*see below*). *The following estimates hold true*:

$$W_1=O\left(\tau\exp\frac{-2l_0\tau}{|K(a_0)-b^2|}\right)\ \ as\ \ \tau\to+\infty,$$
$$W_1=O\left(\exp\frac{-(l_1+b)\tau}{|K(a_1)-b^2|}\right)\ \ as\ \ \tau\to-\infty.$$

Proof. The main equations, obtained by substituting solution (3.6.1.6) into Eq. (3.6.1.1), have the form (3.6.1.10), (3.6.1.11).

The function $\chi(\tau)$ is the solution to the ordinary differential equation

$$b\frac{d\chi}{d\tau}+\frac{d}{d\tau}\left(|K(\chi)-b^2|\frac{d\chi}{d\tau}\right)-F(\chi)=0, \tag{3.6.2.4}$$
$$\chi|_{\tau\to+\infty}=a_0,\quad \chi|_{\tau\to-\infty}=a_1.$$

Note that in this equation there is a "plus" in front of the second derivative. The functions $\beta(t)$ and $\varphi(t)$ satisfy system (3.6.2.1), and for $Lu=O(\varepsilon^2)$ to be valid it is sufficient that Eq. (3.6.1.11) be valid and that the boundary conditions

$$W_1|_{\tau\to-\infty}=0,\quad W_1|_{\tau\to+\infty}=0$$

be met.

The general solution to Eq. (3.6.1.11) has the form

$$W_1=C_1\frac{d\chi}{d\tau}+C_2\frac{d\chi}{d\tau}\int\limits_{-\infty}^{\tau}V\left(\frac{d\chi}{d\tau}\right)^{-1}d\tau \tag{3.6.2.5}$$
$$-\frac{1}{\beta^2\lambda(-\varphi,t)}\frac{d\chi}{d\tau}\int\limits_{+\infty}^{\tau}V\left(\frac{d\chi}{d\eta}\right)^{-2}\left(\int\limits_{+\infty}^{\eta}\frac{f(d\chi/d\tau)\,d\tau}{V(K(\chi)-b^2)}\right)d\eta.$$

The function V, or the Wronskian of Eq. (3.6.1.11), is

$$V=(K(\chi)-b^2)^{-2}\exp\left(b\int^{\tau}\frac{d\tau}{K(\chi)-b^2}\right). \tag{3.6.2.6}$$

Equation (3.6.2.4), as shown in Section 3.2, is related to the equation

$$b\frac{d\Theta}{d\xi}-\frac{d^2\Theta}{d\xi^2}-R(\Theta)=0,\quad R=F(\Theta)(K(\Theta)-b^2)<0,$$

with $\Theta\in(a_0, a_1)$, and has a monotone solution, $d\Theta/d\xi\leqslant 0$, that satisfies the conditions

$$\Theta|_{\xi\to+\infty}=a_1,\quad \Theta|_{\xi\to-\infty}=a_0.$$

This equation can be studied by the same methods as were applied in Section 3.2 (the Zeldovich equation with a nonpositive sink function).

Here the constant b is negative (see Section 3.2) and the following estimates hold true:

$$\Theta\sim a_1-\exp(-l_1\xi)\ \text{as}\ \xi\to+\infty,$$
$$l_1=-b/2+\sqrt{b^2/4+(dF(\Theta)/d\Theta|_{\Theta=a_1})|K(a_1)-b^2|}>0;$$
$$\Theta\sim a_0+\exp(l_0\xi)\ \text{as}\ \xi\to-\infty,$$
$$l_0=b/2+\sqrt{b^2/4+(dF(\Theta)/d\Theta|_{\Theta=a_0})|K(a_0)-b^2|}>0.$$

There exists an inequality that determines the maximal speed:

$$b>b_{\max}=-2\left(\frac{|dF(\Theta)/d\Theta|_{\Theta=a_1}|K(a_1)}{4|dF(\Theta)/d\Theta|_{\Theta=a_1}|-1}\right)^{1/2}.$$

In view of the results arrived at in Section 3.2, the function χ has the asymptotic form, as $\tau\to-\infty$,

$$\chi\sim a_1-\exp\frac{l_1\tau}{|K(a_1)-b^2|}-\exp\frac{2l_1\tau}{|K(a_1)-b^2|}-\dots \tag{3.6.2.7}$$

and the following estimate holds true:

$$\chi\sim a_0+\exp\left(-\frac{l_0\tau}{|K(a_1)-b^2|}\right)+\dots\ \text{as}\ \tau\to+\infty.$$

The function χ is monotone, $d\chi/d\tau<0$.

The following estimates hold true:

$$f=O\left(\tau\exp\frac{-l_0\tau}{|K(a_1)-b^2|}\right)\ \text{as}\ \tau\to+\infty,$$
$$V=O\left(\exp\frac{-b\tau}{|K(a_1)-b^2|}\right)\ \text{as}\ \tau\to+\infty,$$

and the integrals in (3.6.2.5) are divergent for arbitrary β_1 and S_1.

Nullifying the sum of coefficients of $\exp\frac{-l_0\tau}{|K(a_0)-b^2|}$ as $\tau\to+\infty$, we get the following equation:

$$
\begin{aligned}
&\left[\frac{\partial S_1}{\partial t}-\frac{S_1}{\beta}\left(\frac{d\beta}{dt}+2\beta_1\frac{d\varphi}{dt}\right)\right]\\
&-\frac{l_0}{|K(a_0)-b^2|}\left[2\beta\frac{d\varphi}{dt}\frac{\partial S_1}{\partial t}-2S_1\frac{d\varphi}{dt}\left(\frac{d\beta}{dt}+2\beta_1\frac{d\varphi}{dt}\right)\right]\\
&+\left[\frac{d}{dt}\left(\beta\frac{d\varphi}{dt}\right)+2\beta_1\left(\frac{d\varphi}{dt}\right)^2\right]-\beta\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi}\frac{K(a_0)\,l_0}{|K(a_0)-b^2|}\\
&-\beta\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi}K(a_0)-2\beta_1\lambda(-\varphi,\ t)\,K(a_0)\\
&-4\beta_1\lambda(-\varphi,\ t)\,S_1K(a_0)\frac{l_0}{|K(a_0)-b^2|}\\
&+2\lambda(-\varphi,\ t)\,K(a_0)\,\beta\frac{\partial S_1}{\partial x}\frac{l_0}{|K(a_0)-b^2|}=0.
\end{aligned}
\tag{3.6.2.8}
$$

Nullifying the sum of coefficients of $\tau\exp\frac{-l_0\tau}{|K(a_0)-b^2|}$, we get an equation for determining $\beta_1(t)$:

$$
\begin{aligned}
&\beta^{-1}\left(\frac{d\beta}{dt}+2\beta_1\frac{d\varphi}{dt}\right)-2\frac{l_0}{|K(a_0)-b^2|}\frac{d\varphi}{dt}\left(\frac{d\beta}{dt}+2\beta_1\frac{d\varphi}{dt}\right)\\
&+\beta\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi}\frac{l_0K(a_0)}{|K(a_0)-b^2|}+4\beta_1\lambda(-\varphi,\ t)\frac{l_0K(a_0)}{|K(a_0)-b^2|}=0.
\end{aligned}
\tag{3.6.2.9}
$$

From this follows formula (3.6.2.3′) for function β_1. Combining Eqs. (3.6.2.8) and (3.6.2.9), we obtain an equation for determining S_1:

$$
\begin{aligned}
&\frac{\partial S_1}{\partial t}-\frac{2l_0\beta}{|K(a_0)-b^2|}\frac{d\varphi}{dt}\frac{\partial S_1}{\partial t}+\left[\frac{d}{dt}\left(\beta\frac{d\varphi}{dt}\right)+2\beta_1\left(\frac{d\varphi}{dt}\right)^2\right]\\
&-\beta\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi}K(a_0)+2\lambda(-\varphi,\ t)\frac{l_0K(a_0)}{|K(a_0)-b^2|}\beta\frac{\partial S_1}{\partial x}=0.
\end{aligned}
\tag{3.6.2.10}
$$

Let us consider the denominator in (3.6.2.3″) (see also (3.6.1.15)). If we use (3.6.2.3′), we get

$$
\begin{aligned}
&2\beta\frac{d\varphi}{dt}-4\frac{l_0}{|K(a_0)-b^2|}\left(\frac{d\varphi}{dt}\right)^2+4\lambda(-\varphi,\ t)\frac{l_0K(a_0)}{|K(a_0)-b^2|}\\
&=2\lambda(-\varphi,\ t)\left[b-2\frac{l_0b^2}{|K(a_0)-b^2|}+2\frac{l_0K(a_0)}{|K(a_0)-b^2|}\right]\\
&=2\lambda(-\varphi,\ t)\,[b-2l_0]<0.
\end{aligned}
$$

We see that the denominator does not vanish.

Let us continue with the study of (3.6.2.5). Equation (3.6.1.11) implies that if Eqs. (3.6.2.8) and (3.6.2.9) are satisfied, the function f has the following estimate:

$$f = O\left(\tau \exp \frac{-2l_0\tau}{|K(a_0) - b^2|}\right) \text{ as } \tau \to +\infty.$$

The following estimatєs also hold true:

$$fV^{-1}\left(\frac{d\chi}{d\tau}\right) = O\left(\tau \exp \frac{(b-3l_0)\tau}{|K(a_0)-b^2|}\right) \text{ as } \tau \to +\infty,$$

$$V\left(\frac{d\chi}{d\tau}\right)^{-2} = O\left(\exp \frac{(2l_0 - b)\tau}{|K(a_0)-b^2|}\right) \text{ as } \tau \to +\infty.$$

The inner integral in (3.6.2.5) has the estimate

$$\int_{+\infty}^{\tau} fV^{-1}\left(\frac{d\chi}{d\tau}\right) d\tau = O\left(\tau \exp \frac{(b-3l_0)\tau}{|K(a_0)-b^2|}\right) \text{ as } \tau \to +\infty$$

and is convergent, since $b - 3l_0 < 0$.

The outer integral in (3.6.2.5) has the estimate

$$I_1 = \int_{+\infty}^{\tau} V\left(\frac{d\chi}{d\tau'}\right)^{-2}\left(\int_{+\infty}^{\tau'} f\frac{d\chi}{d\tau} V^{-1}\, d\tau\right) d\tau'$$

$$= O\left(\tau \exp \frac{-l_0\tau}{|K(a_0)-b^2|}\right).$$

Thus, the following estimate holds true:

$$W_1 = O\left(\tau \exp \frac{-2l_0\tau}{|K(a_0)-b^2|}\right) \text{ as } \tau \to +\infty.$$

Let us study the behavior of W_1 as $\tau \to -\infty$. By virtue of estimate (3.6.2.7), we have

$$f = O\left(\tau \exp \frac{l_1\tau}{|K(a_1)-b^2|}\right) \text{ as } \tau \to -\infty,$$

and the following estimates hold true:

$$f\left(\frac{d\chi}{d\tau}\right)V^{-1} = O\left(\tau \exp \frac{(b+2l_1)\tau}{|K(a_1)-b^2|}\right) \text{ as } \tau \to -\infty,$$

$$V\left(\frac{d\chi}{d\tau}\right)^{-2} = O\left(\exp \frac{-(b+2l_1)\tau}{|K(a_1)-b^2|}\right) \text{ as } \tau \to -\infty.$$

Since $b + 2l_1 > 0$, the inner integral in (3.6.2.5) is convergent

$$\int_{+\infty}^{-\infty} fV^{-1}\left(\frac{d\chi}{d\tau}\right) d\tau = \text{const}, \quad C_2 \equiv 0,$$

while for the outer integral we have the estimate

$$I_1 = O\left(\exp \frac{-(b+2l_1)\tau}{|K(a_1)-b^2|}\right) \text{ as } \tau \to -\infty.$$

Thus, we have the following estimate

$$W_1 = O\left(\exp \frac{-(b+l_1)\tau}{|K(a_1)-b^2|}\right) \text{ as } \tau \to -\infty$$

since $-b - l_1 > 0$ because $b < 0$.

We have just constructed an asymptotic solution to problem (3.6.1.1), (3.6.1.2) for $K(\chi) - b^2 < 0$. The proof of Theorem 3.6.2.1 is complete.

The results of Theorems 3.6.1.1 and 3.6.2.1 suggest the following

Remark 3.6.2.1 The following estimate holds true:

$$W_1/W_0 = o(1) \text{ as } |\tau| \to \infty$$

if $C_1 \equiv 0$.

3.6.3 KPP Waves in Nonhomogeneous Media

In this section we give formulas describing the propagation of Kolmogorov-Petrovskii-Piskunov (KPP) waves in nonhomogeneous media with slowly varying properties. It appears that in the class of smooth perturbations a wave with the minimal speed is unstable, contrary to the result obtained in Sections 3.6.1 and 3.6.2.

3.6.3.1 An Asymptotic Solution to the Semi-linear Parabolic Equation with Constant Roots in the Equation $F(u) = 0$

The wave solutions to the KPP equation (see Section 3.2.1) have been studied in the classical work of Kolmogorov, Petrovskii, and Piskunov [3.26]. The equation is

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} - R(u) = 0. \tag{3.6.3.1}$$

The invariant solutions to this equation are of the form $u(x, t) = \Theta(x + bt)$, where the function $\Theta = \Theta(\xi)$ is the solution to the following boundary value problem (we assume that $R(0) = 0$, $R(1) = 0$, $R(u) > 0$ for $u \in (0, 1)$):

$$\begin{gathered} b\frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} - R(\Theta) = 0, \\ \lim_{\xi \to -\infty} \Theta(\xi) = 0, \ \lim_{\xi \to +\infty} \Theta(\xi) = 1. \end{gathered} \tag{3.6.3.2}$$

This boundary value problem has a solution if $b \geqslant b_{\min} = 2\,(dR/d\Theta\,|_{\Theta=0})^{1/2}$ (see Section 3.2.1, where we set forth the results of [3.25, 3.26]). An interesting property of the KPP equation is the tendency of the solution to the Cauchy problem to become,

as $t \to \infty$, an invariant solution corresponding to $b = b_{\min}$. In other words, the solution to the KPP equation, $u(x, t)$, such that $u(x, 0) = a(x)$, with $0 \leqslant a(x) \leqslant 1$, $a(x) \equiv 1$, $x \gg 1$, $a(x) \equiv 0$, $x \ll 0$, and $da/dx \geqslant 0$, possesses the following estimate [3.25-3.30][7]

$$u(x, t) = \Theta(x + tb_{\min}) - \delta(x, t),$$
$$\| \delta(x, t) \|_{C(R^1_x)} \to 0, \text{ as } t \to \infty.$$

In some papers this property is called the stability of a simple wave with the minimal speed.

Let us consider the KPP equation with variable coefficients and a small parameter acting as a coefficient of the derivatives:

$$Lu = \varepsilon \frac{\partial u}{\partial t} - \varepsilon^2 \frac{\partial}{\partial x} \left(\lambda(x, t) \frac{\partial u}{\partial x} \right) - \gamma^2 (x, t) R(u) = 0, \quad (3.6.3.3)$$
$$u|_{x \to \infty} = 1, \quad u|_{x \to -\infty} \to 0,$$

where $\lambda(x, t)$ and $\gamma(x, t)$ are smooth positive functions. Parameter ε in the equation emerges, for instance, if the solution to the initial KPP equation (3.6.3.1) is considered for large times $\tilde{t} \sim t/\varepsilon$ and "large" $\tilde{x} \sim x/\varepsilon$. (Going over to variables $\tilde{x}$ and $\tilde{t}$ in Eq. (3.6.3.1) and canceling out the wave, we arrive at Eq. (3.6.3.1) with $\lambda \equiv 1$ and $\gamma \equiv 1$.) Below we construct asymptotic solutions to Eq. (3.6.3.1) of the form

$$u(x, t) = \Theta(S/\varepsilon) + \varepsilon W_1 (S/\varepsilon, t, \varepsilon), \quad (3.6.3.4)$$

where

$$S(x, t) = \beta(t)(x + \varphi) + \beta_1(t)(x + \varphi^2) + \varepsilon S_1(x, t).$$

Essentially these solutions are distorted simple waves, and the existence of such waves means that the KPP waves are stable under slow variations of the properties of the external medium.

It appears, however, that the solution to Eq. (3.6.3.1) of the form (3.6.3.4) exists only for $b > b_{\min}$ and, hence, the wave with the minimal speed proves to be unstable under such variations of the properties of the medium.

On the other hand, if all the terms in the law of energy conservation are taken into account, that is, if the term $\partial^2 u/\partial t^2$ is retained, the result is a stable wave in relation to this class of medium perturbations.

In Sections 3.6.1 and 3.6.2 we gave the result of the investigation of the complete equation (it can be assumed that $K(u) \equiv 1$). Below

---

[7] Provided that $a(\xi)$ is not a solution to Eq. (3.6.2.2) for any value $b > b_{\min}$. In [3.25] there are references to the works of R.A. Fisher, Y.A. Kanal, P.G. Fife, J.B. McLeog, D.G. Aronson, H.F. Weinberger, and others (see also [3.27-3.29]).

we employ this result in connection with Eq. (3.6.3.3). Let us give a formal solution to problem (3.6.3.3). We have

Theorem 3.6.3.1 *Let* $R(u) > 0$, $u \in (0, 1)$. *Then an asymptotic solution of the* (3.6.3.4) *type to problem* (3.6.3.3) *exists, and the function* Θ *is a solution to problem* (3.6.3.2) *for* $b > b_{\min}$. *The functions* $\beta(t)$ *and* $\varphi(t)$ *are defined through the system of equations*

$$\beta^2\lambda(-\varphi, t) = \gamma(-\varphi(t), t), \ \beta\frac{d\varphi}{dt} = \gamma(-\varphi(t), t)\, b, \tag{3.6.3.5}$$

and the function $\beta_1(t)$ *is defined thus:*

$$\beta_1(t) = \left[\beta\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi} l^2 - \frac{d\beta}{dt}\frac{l}{\beta} + \left.\frac{\partial\gamma}{\partial x}\right|_{x=-\varphi}\beta^{-1}\right] \times \left[2\lambda(-\varphi, t)\, l\sqrt{b^2 - b_{\min}^2}\right]^{-1} \tag{3.6.3.6}$$

where

$$b_{\min} = 2\sqrt{dR/d\Theta|_{\Theta=0}}, \quad l = (b - \sqrt{b^2 - b_{\min}^2})/2.$$

The function S_1 *can be found by solving the equation*

$$-\frac{\partial S_1}{\partial t} + 2\lambda(-\varphi, t)\beta l\frac{\partial S_1}{\partial x} + \beta\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi} + 2\beta_1\lambda(-\varphi, t) = 0.$$

The function W_1 *has the form*

$$W_1(\xi, t) = C_1\frac{d\Theta}{d\xi} - \lambda^{-1}(-\varphi, t)\beta^{-2}\frac{d\Theta}{d\xi} \times \int_{-\infty}^{\xi} V\left(\frac{d\Theta}{d\xi'}\right)^{-2}\left(\int_{\infty}^{\xi'} fV^{-1}\frac{d\Theta}{d\mu}\, d\mu\right) d\xi' \tag{3.6.3.7}$$

and the following estimates hold true (see the Remark on p. 320):

$$W_1 = O(\xi\exp(2l\xi)) \text{ as } \xi \to -\infty, \ b > \frac{3}{\sqrt{8}} b_{\min};$$

$$W_1 = O(\xi^2\exp((b-l)\xi)) \ \textit{as} \ \xi \to -\infty, \ b = \frac{3}{\sqrt{8}} b_{\min}, \quad b - l > l;$$

$$W_1 = O(\exp((b-l)\xi)) \ \textit{as} \ \xi \to -\infty, \ b_{\min} < b < \frac{3}{\sqrt{8}} b_{\min},$$

$$W_1 = O(\xi^2\exp(-l_0\xi)) \ \textit{as} \ \xi \to \infty.$$

Remark 3.6.3.1 An asymptotic solution to problem (3.6.3.3) is generally described by formulas given in Theorem 3.6.3.1 only for the values of variables x and t that obey the condition $\beta(x + \varphi) + \beta_1(x + \varphi)^2 \sim (x + \varphi)$. Outside this region the solution is continued by zero and unity, respectively, just as in Section 3.4.

Proof. Let us substitute the function (3.6.3.4) into Eq. (3.6.3.3). As in Section 3.4.1, we obtain

$$Lu=\left\{\beta\frac{d\varphi}{dt}\frac{d\Theta}{d\xi}-\beta^2\lambda(-\varphi,\ t)\frac{d^2\Theta}{d\xi^2}-\gamma(-\varphi,\ t)R(\Theta)\right\}$$

$$+\varepsilon\left\{\beta\frac{d\varphi}{dt}\frac{\partial W_1}{\partial\xi}-\beta^2\lambda(-\varphi,\ t)\frac{\partial^2 W_1}{\partial\xi^2}-\gamma(-\varphi,\ t)\frac{dR(\Theta)}{d\Theta}W_1\right\}$$

$$+\varepsilon\Big[-\Big\{\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}(\xi-S_1)\frac{d^2\Theta}{d\xi^2}+\beta\frac{d\Theta}{d\xi}\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}$$

$$+2\beta_1\lambda(-\varphi,\ t)\frac{d\Theta}{d\xi}+4\beta_1(-\varphi,\ t)(\xi-S_1)\frac{d^2\Theta}{d\xi^2}$$

$$-\frac{\partial\Theta}{\partial t}-\left(2\beta_1\frac{d\varphi}{dt}+\frac{d\beta}{dt}\right)\frac{\xi-S_1}{\beta}\frac{d\Theta}{d\xi}+\frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi}\frac{\xi-S_1}{\beta}R(\Theta)\Big\}$$

$$+\varepsilon\left\{\frac{d\Theta}{d\xi}\frac{\partial S_1}{\partial t}-\lambda(-\varphi,\ t)\frac{d^2\Theta}{d\xi^2}2\beta-\frac{\partial S_1}{\partial x}\right\}\Big]+O(\varepsilon^2).\quad(3.6.3.8)$$

If $\beta(t)$ and $\varphi(t)$ are found by solving system (3.6.3.5) and Θ is the solution to problem (3.6.3.2), then $Lu=O(\varepsilon)$. For the condition $Lu=O(\varepsilon)$ to be satisfied, it is sufficient that the following be valid:

$$\beta\frac{d\varphi}{dt}\frac{\partial W_1}{\partial\xi}-\beta^2\lambda(-\varphi,\ t)\frac{\partial^2W_1}{\partial\xi^2}-\gamma(-\varphi,\ t)\frac{dR(\Theta)}{d\Theta}W_1=f(t,\ \varepsilon),\tag{3.6.3.9}$$

where f stands for the sum inside the square brackets in (3.6.3.8).

The boundary conditions (3.6.3.3) imply the following boundary conditions for W_1:

$$W_1|_{\xi\to-\infty}=0,\quad W_1|_{\xi\to+\infty}=0.$$

The general solution to Eq. (3.6.3.9) has the form

$$W_1=C_1\frac{d\Theta}{d\xi}+C_2\frac{d\Theta}{d\xi}\int_{-\infty}^{\xi}V\left(\frac{d\Theta}{d\xi'}\right)^{-2}d\xi'$$

$$+\lambda^{-1}(-\varphi,\ t)\beta^{-2}\frac{d\Theta}{d\xi}\int_{-\infty}^{\xi}V\left(\frac{d\Theta}{d\xi'}\right)^{-2}\left(\int_{\infty}^{\xi'}fV^{-1}\frac{d\Theta}{d\xi}d\xi\right)d\xi',\tag{3.6.3.10}$$

with $V=V(\xi)$ the Wronskian of Eq. (3.6.2.9),

$V=\exp(b\xi)$.

As $\xi\to-\infty$ the following estimate holds true (see Section 3.2):

$$\Theta(\xi)\sim\exp(l\xi)+C_3\exp(2l\xi)+\ldots,\ l=(b-\sqrt{b^2-b_{\min}^2})/2.\tag{3.6.3.11}$$

Hence, as $\xi \to -\infty$, we have the estimates

$$f = O(\xi \exp(l\xi)) + O(\exp(l\xi)),$$
$$V = O(\exp(b\xi)),$$

and the inner integral in (3.6.3.10) is divergent as $\xi \to -\infty$ for arbitrary β_1 and S_1, since $2l - b < 0$.

Nullifying the sum of coefficients of $\exp(l\xi)$ as $\xi \to -\infty$, we get the equation

$$-l\frac{\partial S_1}{\partial t} + 2\lambda(-\varphi, t)\, l^2\beta \frac{\partial S_1}{\partial x} - \beta \frac{\partial \lambda}{\partial x}\Big|_{x=-\varphi} l^2 S_1 + \beta l \frac{\partial \lambda}{\partial x}\Big|_{x=-\varphi}$$
$$+ 2\beta_1\lambda(-\varphi, t)\, l - 4\beta_1\lambda(-\varphi, t)\, S_1 l^2$$
$$+ \left[2\beta_1 \frac{d\varphi}{dt} + \frac{d\beta}{dt}\right]\frac{lS_1}{\beta} - \frac{S_1}{\beta}\frac{\partial \gamma}{\partial x}\Big|_{x=-\varphi} = 0. \tag{3.6.3.12}$$

Nullifying the sum of coefficients of $\xi \exp(l\xi)$ as $\xi \to -\infty$, we get the equation for β_1:

$$\beta l^2 \frac{\partial \lambda}{\partial x}\Big|_{x=-\varphi} + 4\beta_1\lambda(-\varphi, t)\, l^2 - \left[2\beta_1 \frac{d\varphi}{dt} + \frac{d\beta}{dt}\right]\frac{l}{\beta}$$
$$+ \beta^{-1}\frac{\partial \gamma}{\partial x}\Big|_{x=-\varphi} = 0. \tag{3.6.3.13}$$

Combining Eqs. (3.6.3.12) and (3.6.3.13) yields an equation for S_1:

$$-\frac{\partial S_1}{\partial t} + \lambda(-\varphi, t)\, 2\beta l \frac{\partial S_1}{\partial x} + \beta l \frac{\partial \lambda}{\partial x}\Big|_{x=-\varphi}$$
$$+ 2\beta_1\lambda(-\varphi, t) = 0. \tag{3.6.3.14}$$

The solution to Eq. (3.6.3.13) has the form

$$\beta_1(t) = -\left[l^2\beta \frac{\partial \lambda}{\partial x}\Big|_{x=-\varphi} - \frac{l}{\beta}\frac{d\beta}{dt} + \beta^{-1}\frac{\partial \gamma}{\partial x}\Big|_{x=-\varphi}\right]$$
$$\times \left[4\lambda(-\varphi, t)\, l^2 - 2\frac{l}{\beta}\frac{d\varphi}{dt}\right]^{-1}. \tag{3.6.3.15}$$

Let us calculate the denominator in (3.6.3.15). Substituting $d\varphi/dt$ and β from system (3.6.3.6) and the value of l from the hypothesis of Theorem 3.6.3.1, we obtain

$$-4\lambda(-\varphi, t)\, l^2 + \frac{2l}{\beta}\frac{d\varphi}{dt} = 2\lambda(-\varphi, t)\, l \sqrt{b^2 - b_{\min}^2} > 0.$$

Let us continue with the study of (3.6.3.10). If Eqs. (3.6.3.13) and (3.6.3.14) are valid, the function f has the estimate

$$f = O(\xi \exp 2l\xi) \text{ as } \xi \to -\infty. \tag{3.6.3.16}$$

Moreover, there are the following estimates (as $\xi \to -\infty$):

$$fV^{-1}\frac{d\Theta}{d\xi}=O(\xi \exp((3l-b)\xi)),$$
$$V\left(\frac{d\Theta}{d\xi}\right)^{-2}=O(\exp((b-2l)\xi)),\quad b-2l>0. \tag{3.6.3.17}$$

Let us analyze the exponents as $\tau \to -\infty$:

$$b-2l=\sqrt{b^2-b_{\min}^2}>0,$$
$$3l-b=b/2-(3/2)\sqrt{b^2-b_{\min}^2}>0 \tag{3.6.3.18}$$

for $b_{\min}<b<(3/\sqrt{8})\,b_{\min}$.

Thus, the following estimate holds true for the inner integral in (3.6.3.10):

$$I=\int_{-\infty}^{\xi} f\frac{d\Theta}{d\xi}V^{-1}\,d\xi=O(\tau\exp((3l-b)\tau)),$$

and for $b_{\min}<b<(3/\sqrt{8})_{\min}$ the integral is convergent as $\xi \to -\infty$, while for $b \geqslant (3/\sqrt{8})\,b_{\min}$ it is divergent as $\xi \to -\infty$. The following estimate holds true:

$$I=O(\xi^2),\quad b=(3/\sqrt{8})\,b_{\min}.$$

As $\tau \to -\infty$ the following estimates hold true:

$$I_0=\int_{-\infty}^{\xi} V\frac{d\Theta}{d\xi'}\left(\int_{+\infty}^{\xi'} fV^{-1}\frac{d\Theta}{d\xi}\,d\xi\right)d\xi'$$
$$=O(\exp((b-2l)\xi)),\quad b-2l>0,$$
$$b_{\min}<b<(3/\sqrt{8})\,b_{\min};$$
$$I_0=O(\xi\exp(l\xi)),\quad b>(3/\sqrt{8})\,b_{\min};$$
$$I_0=O(\xi^2\exp((b-2l)\xi)),\quad b=(3/\sqrt{8})\,b_{\min}.$$

Thus, $C_2\equiv 0$ and (3.6.3.7) holds true as $\tau \to -\infty$. This implies that the following estimates hold true:

$$W_1=O(\exp((b-l)\xi)),\quad b_{\min}<b<(3/\sqrt{8})\,b_{\min};$$
$$W_1=O(\xi^2\exp((b-l)\xi),\quad b=(3/\sqrt{8})\,b_{\min},\quad b-l>l; \tag{3.6.3.19}$$
$$W_1=O(\xi\exp(2l\xi)),\quad b>3/\sqrt{8})\,b_{\min}.$$

Let us consider the behavior of the integrand in (3.6.3.7) as $\xi \to \infty$. By virtue of the estimate (see Section 3.2)

$$\Theta \sim 1-\exp(-l_0\xi),\quad l_0=-b/2+\sqrt{b^2/4+|dR/d\Theta|_{\Theta=1}|}>0,$$

we have (as $\xi \to \infty$)

$$\int\limits_{\infty}^{\tau} fV^{-1} \frac{d\Theta}{d\xi}\, d\xi = O\left(\xi \exp\left((2l_0 - b)\,\xi\right)\right),$$
$$V\left(\frac{d\Theta}{d\xi}\right)^{-2} = O\left(\exp\left(b + 2l_0\right)\xi\right)). \tag{3.6.3.20}$$

In view of the first estimate in (3.6.3.20), the outer integral in (3.6.3.7) is convergent as $\xi \to \infty$. This formula yields the estimate

$$W_1 = O\left(\xi^2 \exp\left(-l_0\xi\right)\right) \quad \text{as} \quad \xi \to \infty.$$

The proof of the theorem is complete.

Let us investigate the condition $b > b_{\min}$ used in the proof of the theorem. If $b = b_{\min}$ the denominator in (3.6.3.16) vanishes and, hence, the equation for finding $\beta_1(t)$ has no solution and the function $fV^{-1}(d\Theta/d\xi)$ grows in direct proportion to $|\xi|$ as $\xi \to -\infty$. Thus, at $b = b_{\min}$ the integral in (3.6.3.18) has no finite value for $\xi' < \infty$ and, as can easily be seen, the equation for finding W_1 has no useful solution (i.e. a solution that decreases as $|\xi| \to \infty$). One could attempt to construct the next term in the asymptotic expansion of the solution by employing the method of the "operator-valued symbol" [3.3], that is, by writing a partial differential equation for W_1 (retaining the derivatives $\varepsilon\,(\partial W_1/\partial t)$ and $\varepsilon\,(\partial W_1/\partial x)$), which is a linear equation with variable coefficients:

$$\varepsilon \frac{\partial W_1}{\partial t} = A\left(\frac{\partial}{\partial \tau},\ \varepsilon \frac{\partial}{\partial x}\right) W_1 + f.$$

The operator-valued symbol $A\,(\partial/\partial\tau, p)$ in this case proves to be nonhermitian (because of the presence of $\partial W_1/\partial\tau$), but a unitary transformation can be applied to reduce this operator to a hermitian nonpositive operator in the $L_2(R^1)$ space with a weight function $\exp(-b\xi)$. However, the fact that the equation for $\beta_1(t)$ has no solution at $b = b_{\min}$ means that the new right-hand side does not belong to $L_2(R^1)$ with the weight function $\exp(-b\xi)$.

The propagation of a distorted KPP wave at $b = b_{\min}$ is possible only if $\lambda = \lambda(t)$, $\gamma = \gamma(t)$, and $\lambda/\gamma = \text{const}$. Then Eq. (3.6.3.13) becomes an identity for every $\beta_1(t)$. Note that in this case Eq. (3.6.3.3) in variables x and $\tilde{t} = \int\limits_0^t \lambda(t)dt$ is an equation with constant coefficients and possesses an exact invariant solution of the wave type. There are also other ways of constructing the second term in the asymptotic solution, but in Theorem 3.6.3.1 we have employed the method in which the following estimates hold true:

$$W_1/W_0 = o(1) \text{ as } \xi \to -\infty \text{ if } C_1 = 0,$$
$$W_1/W_0 = o(1) \text{ as } \xi \to \infty \text{ since } \lim_{\xi\to\infty} W_0 = 1 - 0.$$

3.6.3.2 An Asymptotic Solution to the Semi-linear Equation with a Variable Root of the Equation $F = 0$

In this section we provide a solution to the problem of a parabolic equation of the (3.6.3.1) type with a variable root of the equation $F(u, x, t) = 0$. The problem has the form

$$\varepsilon \frac{\partial u}{\partial t} - \varepsilon^2 \frac{\partial}{\partial x}\left(\lambda(x, t)\frac{\partial u}{\partial x}\right) - \gamma(x, t)\, u\,(\mu(x) - u) = 0, \tag{3.6.3.21}$$

$$u(-\infty, t) = 0, \quad u(\infty, t) = \mu(x, t).$$

The asymptotic solution to problem (3.6.3.21) will be sought in the form

$$u(x, t, \varepsilon) = \mu(x)\,[\Theta(S/\varepsilon) + \varepsilon W_1(S/\varepsilon, t)], \tag{3.6.3.22}$$

where the function S has the form

$$S(x, t, \varepsilon) = \beta(t)(x + \varphi(t)) + \beta_1(t)(x + \varphi(t))^2 + \varepsilon S_1(x, t).$$

If $\mu = \mu(x, t)$, the algorithm for constructing the solution is given in [3.3].

Theorem 3.6.3.2 *Problem* (3.6.3.21) *has an asymptotic solution of the* (3.6.3.22) *type. The function* $\Theta(\xi)$ *is the solution to*

$$b\frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} - \Theta(1 - \Theta) = 0, \qquad \Theta(-\infty) = 0, \quad \Theta(\infty) = 1. \tag{3.6.3.23}$$

The functions β *and* φ *can be found by solving the system of equations*

$$\beta\frac{d\varphi}{dt} = b\gamma(-\varphi, t)\,\mu(-\varphi), \quad \lambda(-\varphi, t)\,\beta^2 = \gamma(-\varphi, t)\,\mu(-\varphi). \tag{3.6.3.24}$$

The function β_1 *has the form*

$$\beta_1 = -\left[\beta l^2 \frac{\partial(\lambda\mu)}{\partial x}\bigg|_{x=-\varphi} - \mu(-\varphi)\frac{d\beta}{dt}\frac{l}{\beta} + \beta^{-1}\frac{\partial(\gamma\mu^2)}{\partial x}\bigg|_{x=-\varphi} - l\frac{d\varphi}{dt}\frac{\partial\mu}{\partial x}\bigg|_{x=-\varphi}\right] \times [2\lambda(-\varphi, t)\,\mu(-\varphi)B]^{-1},$$

$$B = l - \sqrt{b^2 - b_{\min}^2}, \quad l = (b - \sqrt{b^2 - 4})/2, \quad b_{\min} = 2. \tag{3.6.3.24'}$$

The function $S_1(x, t)$ can be found by solving the equation

$$\frac{\partial S_1}{\partial t}\mu(-\varphi)l-\lambda(-\varphi, t)\mu(-\varphi)l^2 2\beta\frac{\partial S_1}{\partial x}$$
$$-\mu(-\varphi)l\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}-\lambda(-\varphi, t)2l\beta\frac{\partial\mu}{\partial x}\Big|_{x=-\varphi}$$
$$-2\beta_1\lambda(-\varphi, t)\mu(-\varphi)l=0.$$

The function W_1 has the form

$$W_1(\xi, t)=C_1\frac{d\Theta}{d\xi}-\lambda^{-1}(-\varphi, t)\beta^{-2}(t)\frac{d\Theta}{d\xi}$$
$$\times\int_{-\infty}^{\xi}V\left(\frac{d\Theta}{d\xi'}\right)^{-2}\left(\int_{+\infty}^{\xi'}fV^{-1}\frac{d\Theta}{d\mu}d\mu\right)d\xi', \quad (3.6.3.25)$$

with

$$f=-\Big\{\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}(\xi-S_1)\frac{d^2\Theta}{d\xi^2}+\beta\frac{d\Theta}{d\xi}\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}$$
$$+2\beta_1\lambda(-\varphi, t)\frac{d\Theta}{d\xi}+4\beta_1\lambda(-\varphi, t)(\xi-S_1)\frac{d^2\Theta}{d\xi^2}$$
$$-\left(2\beta_1\frac{d\varphi}{dt}+\frac{d\beta}{dt}\right)\frac{\xi-S_1}{\beta}\frac{d\Theta}{d\xi}+\frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi}\frac{\xi-S_1}{\beta}F(\Theta)\Big\}$$
$$+\Big\{\left(\frac{\partial S_1}{\partial t}+\beta_3\frac{d\varphi}{dt}\right)\frac{d\Theta}{d\xi}\lambda(-\varphi, t)\frac{d^2\Theta}{d\xi^2}2\beta\frac{\partial S_1}{\partial x}\Big\},$$
$$V=\exp(b\xi).$$

Proof. The proof of this theorem is similar to the proof of Theorem 3.6.3.1. Obviously, for $\mu\equiv 1$ Eqs. (3.6.3.24) and (3.6.3.25) transform into the corresponding formulas in Theorem 3.6.3.1. At $b=b_{\min}$ the denominator in (3.6.3.24′) vanishes, that is, in this case there is no asymptotic solution of the type considered.

3.6.4 The Zeldovich and Semyonov Waves in Nonhomogeneous Media[8]

In this section we study the asymptotic solutions to the Zeldovich and Semyonov equations. We will demonstrate that, in contrast to the above solution to the KPP equation, these solutions are always stable in the class of smooth perturbations.

3.6.4.1 The Zeldovich Equation

The theory of propagation of the laminar front of flame (plasma-gas) widely employs a model first proposed by Ya. B. Zeldovich. In this section we briefly discuss an algorithm for constructing

[8] In the English-language literature the Semyonov equation is sometimes called the Fitzhugh-Nagumo equation.

an asymptotic solution describing the diffusion-thermal structure of the flame front in a nonhomogeneous medium. Within the framework of the Zeldovich model, the propagation of a flame front is described by the following semi-linear equation:

$$\varepsilon \frac{\partial u}{\partial t} - \varepsilon^2 \frac{\partial}{\partial x}\left(\lambda(x, t)\frac{\partial u}{\partial x}\right) - \gamma(x, t) R(u) = 0, \qquad (3.6.4.1)$$
$$x \in R_1, \quad t \in [0, t'], \quad t' > 0.$$

The function $R(u)$ has two roots: $R(0) = 0$ and $R(1) = 0$, with

$$dR/du\,|_{u=0} = 0, dR/du\,|_{u=1} > 0, R > 0 \text{ for } u \in (0, 1). \qquad (3.6.4.2)$$

Let us construct an asymptotic solution to Eq. (3.6.4.1) satisfying the boundary conditions

$$u|_{x\to-\infty} = 0, \quad u|_{x\to+\infty} = 1. \qquad (3.6.4.3)$$

An asymptotic solution to problem (3.6.4.1)-(3.6.4.3) can be constructed by the method developed in Section 3.6.3 and has the form (3.6.3.4). For this reason we give only the result and comments on it.

Theorem 3.6.4.1 *Let the conditions* (3.6.4.2) *be met. Then problem* (3.6.4.1)-(3.6.4.3) *has an asymptotic solution of the form* (3.6.3.4), *with* Θ *the solution to*

$$b_0 \frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} - R(\Theta) = 0,$$
$$\Theta|_{\xi\to-\infty} = 0, \quad \Theta|_{\xi\to\infty} = 1.$$

The function $S(x, t, \varepsilon)$ *has the form*

$$S(x, t, \varepsilon) = (\beta(t) + \varepsilon\beta_2(t))(x + \varphi(t)) + \beta_1(t)(x + \varphi(t))^2 + \varepsilon S_1(t),$$

with b_0 *the Zeldovich constant (see Section* 3.2).

The functions $\beta(t)$ *and* $\varphi(t)$ *can be found by solving the system of equations*

$$\beta^2(t)\lambda(-\varphi, t) = \gamma(-\varphi(t), t), \quad \beta\frac{d\varphi}{dt} = \gamma(-\varphi, t) b_0, \qquad (3.6.4.3')$$

and the function $\beta_1(t)$ *is defined thus:*

$$\beta_1(t) = -\left[\beta \frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi} b_0^2 + \frac{b_0}{\beta}\frac{d\beta}{dt} - \beta^{-1}\frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi}\right] \times [2\lambda(-\varphi, t) b_0^2]^{-1}. \qquad (3.6.4.4)$$

The functions S_1 *and* β_2 *can be found by solving the equations*

$$-\frac{\partial S_1}{\partial t} - \beta_2\frac{d\varphi}{dt} + 2\lambda(-\varphi, t)\beta_2\beta + \beta l\,\frac{d\lambda}{dx}\Big|_{x=-\varphi} + 2\beta_1\lambda(-\varphi, t) = 0, \qquad (3.6.4.5)$$

$$\int_{-\infty}^{\infty} \exp(-b_0\xi)\left\{\left(\frac{\partial S_1}{\partial t}+\beta_2\frac{d\varphi}{dt}\right)\frac{d\Theta}{d\xi}-2\lambda(-\varphi,\ t)\beta\beta_2\frac{d^2\Theta}{d\xi^2}\right.$$
$$-\left[\beta\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi}(\xi-S_1)\frac{d^2\Theta}{d^2\xi}+\beta\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi}\frac{d\Theta}{d\xi}\right.$$
$$-\left(2\beta_1\frac{d\varphi}{dt}+\frac{d\beta}{dt}\right)\frac{1}{\beta}\frac{d\Theta}{d\xi}(\xi-S_1)$$
$$\left.\left.-\beta^{-1}\left.\frac{\partial\gamma}{\partial x}\right|_{x=-\varphi}F(\Theta)(\xi-S_1)\right]\right\}\frac{d\Theta}{d\xi}\,d\xi=0. \qquad (3.6.4.6)$$

The function W_1 has the form

$$W_1=C_1\frac{d\Theta}{d\xi}-\lambda^{-1}(-\varphi,\ t)\beta^{-2}\frac{d\Theta}{d\xi}\int_{-\infty}^{\xi}V\left(\frac{d\Theta}{d\xi'}\right)^{-2}\left(\int_{-\infty}^{\xi'}fV^{-1}\frac{d\Theta}{d\mu}d\mu\right)d\xi', \qquad (3.6.4.7)$$

and the following estimates hold true: [9]

$$W_1=O(\xi\exp(2b_0\xi)) \quad as\ \xi\to-\infty,$$
$$W_1=O(\xi^2\exp(-l_0\xi)) \quad as\ \xi\to\infty. \qquad (3.6.4.8)$$
$$l_0=-b_0/2+\sqrt{b_0^2/4+|dR/d\Theta|_{\Theta=1}|}.$$

Proof. Substituting (3.6.3.4) into Eq. (3.6.4.1), we get (3.6.3.8). The solution of the standard equation is obvious (see Theorem 3.2.2.2).

Let us now analyze problem (3.6.3.9), following the pattern used in Section 3.6.3. Differences appear only in the solution of Eq. (3.6.3.13).

Let us calculate the denominator in (3.6.3.15):

$$4\lambda(-\varphi,\ t)\,l^2-\frac{2l}{\beta}\frac{d\varphi}{dt}=2\lambda(-\varphi,\ t)\,l\,(2l-b_0).$$

In Section 3.2.1 it was found that $b_0 = l$ for equations of the Zeldovich kind, whereby

$$4\lambda(-\varphi,\ t)\,l^2-\frac{2l}{\beta}\frac{d\varphi}{dt}=2\lambda(-\varphi,\ t)\,b_0^2>0.$$

Hence, for equations of the Zeldovich kind the equation for finding β_1 always has a solution. Equation (3.6.4.6) follows from the orthogonality condition

$$\int_{-\infty}^{\infty} fV^{-1}\frac{d\Theta}{d\xi}\,d\xi=0.$$

[9] See Remark 3.6.3.1.

In view of the fact that conditions (3.6.4.4) and (3.6.4.5) are met, the integral in the orthogonality condition has a finite value. The proof of the theorem is complete.

3.6.4.2 The Semyonov Equation

In the mathematical model of autocatalytic chain chemical reactions proposed by N. N. Semyonov there emerges a semilinear parabolic equation with a small parameter ε acting as a coefficient of the derivatives:

$$\begin{aligned}&\varepsilon\frac{\partial u}{\partial t}-\varepsilon^2\frac{\partial}{\partial x}\left(\lambda(x,\,t)\frac{\partial u}{\partial x}\right)-R(u)=0,\\&x\in R^1,\;t\in[0,\,T].\end{aligned}\tag{3.6.4.9}$$

Here, in contrast to the equations considered earlier, the function $R(u)$ has three zeros on the segment $[0,\,1]$:

$$R(0)=0,\;R(a_1)=0,\;R(1)=0,\;a_0\in(0,\,1),$$

and

$$dR/d\Theta|_{\Theta=0}<0,\quad dR/d\Theta|_{\Theta=a_1}>0,\quad dR/d\Theta|_{\Theta=1}<0.\tag{3.6.4.10}$$

Let us construct an asymptotic solution to Eq. (3.6.4.1) satisfying the following conditions:

$$u|_{x\to-\infty}=1,\quad u|_{x\to\infty}=0.\tag{3.6.4.11}$$

Actually an algorithm for solving problem (3.6.4.9)-(3.6.4.11) is given in Section 3.4, so that here we only formulate the result.

Theorem 3.6.4.2 *Suppose that conditions* (3.6.4.10) *are met. Then problem* (3.6.4.9)-(3.6.4.11) *has an asymptotic solution of the* (3.6.3.4) *type. The function* $S(x,\,t,\,\varepsilon)$ *has the form*

$$\begin{aligned}S(x,\,t,\,\varepsilon)&=(\beta(t)+\varepsilon\beta_2(t))(x+\varphi(t))\\&\quad+\beta_1(t)(x+\varphi(t))^2+\varepsilon S_1(x,\,t).\end{aligned}$$

The function Θ *constitutes a solution to the problem*[10]

$$b\frac{d\Theta}{d\xi}-\frac{d^2\Theta}{d\xi^2}-R(\Theta)=0,\quad\Theta|_{\xi\to-\infty}=0,\quad\Theta|_{\xi\to\infty}=1,\tag{3.6.4.12}$$

where $b<2\sqrt{dR/d\Theta_{\Theta=a_1}}$, *and the following estimates hold true*

$$\begin{aligned}&\Theta\sim1-\exp(l\xi)\qquad as\;\xi\to-\infty,\\&\Theta\sim O(\exp(-l_0\xi))\quad as\;\xi\to\infty,\end{aligned}\tag{3.6.4.13}$$

$$\begin{aligned}&l=b/2+\sqrt{b^2/4+|dR/d\Theta|_{\Theta=0}|},\\&l_0=-b/2+\sqrt{b^2/4+|dR/d\Theta|_{\Theta=1}|}.\end{aligned}\tag{3.6.4.14}$$

---

[10] This problem has been studied extensively in Section 3.4.2, see also the Remark on p. 320.

The functions $\beta(t)$ *and* $\varphi(t)$ *can be found by solving the system of equations*

$$\beta^2\lambda(-\varphi,\ t)=\gamma(-\varphi,\ t),\ \beta\frac{d\varphi}{dt}=b\gamma(-\varphi,\ t),$$

and $\beta_1(t)$ *is defined thus*:

$$\beta_1(t)=-\left[\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}l^2-\frac{d\beta}{dt}\frac{l}{\beta}+\frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi}\frac{1}{\beta}\right]$$
$$\times[4\lambda(-\varphi,\ t)\,l\sqrt{b^2/4+|dR/d\Theta|_{\Theta=0}|}]^{-1}.$$

The functions S_1 *and* β_2 *can be found by solving the system of equations* (3.6.4.5) *and* (3.6.4.6). *The function* W_1 *is of the* (3.6.4.7) *type, and the following estimates hold true*:

$$W_1=O(\xi\exp(2l\xi))\qquad as\quad \xi\to-\infty,$$
$$W_1=O(\xi^2\exp(-l_0\xi))\quad as\quad \xi\to+\infty.$$

Proof. The proof is similar to the one used for Theorem 3.6.3.1. We note only that in this case, as in equations of the Zeldovich kind, the denominator in formula (3.6.3.15) never vanishes. Indeed, suffice it to check the inequality $2l-b>0$, which follows from formula (3.6.4.14) for l.

3.6.5 Propagation of Nonlinear Thermal Waves

In this section we consider the problem of a formed thermal wave whose wavefront can release energy (e.g. see [3.34-3.36]). We assume that this energy release is caused by an intensive plasma-chemical reaction proceeding in a definite temperature interval and that the thermal conductivity coefficient decreases as u grows.

The process of propagation of a thermal wave is described by the following equation (in dimensionless form):

$$\varepsilon\frac{\partial u}{\partial t}+\varepsilon^2\frac{\partial^2u}{\partial t^2}-\varepsilon^2\nabla(\lambda(x,\ t)K(u)\nabla u)-F(u)=0, \tag{3.6.5.1}$$

where

$$x\in R^3,\ t\in[0,\ T],\ 0\leqslant\varepsilon<1,\ u\geqslant0,$$
$$K(u)>0,\ F(a_i)=0,\ a_i=\text{const}>0,\ i=0,\ 1,\ a_1>a_0.$$

Here $u=T/T_0$, $x=\bar{x}/x_0$, $\varepsilon=\varkappa_0t_0/x_0^2<1$ is the small parameter in the problem, $t=\bar{t}/(t_0\varepsilon)$, $t_0=x_0/c$, with $\bar{x}$ and $\bar{t}$ the dimensional coordinate and time, T_0 is the characteristic unperturbed temperature of the medium at which heat liberation begins, x_0 is the mean free path of the radiation in the medium, c is the speed of the steady-state thermal wave, and $\varkappa_0$ is the thermal diffusivity of the medium.

The functions $F(u)$, $K(u)$, $K(u)\nabla u$, and $\lambda(x,\ t)$ are continuously differentiable (the first three, for $u>0$), $F(u)$ is the dimen-

sionless source function, $K(u)$ is the dimensionless thermal conductivity of the medium, and $\lambda(x, t) \geqslant \delta = \text{const} > 0$ characterizes the slowly varying properties of the medium. When the internal energy of the gas is proportional to the temperature and the mean free

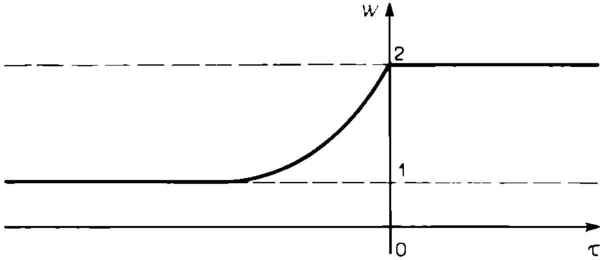


Fig. 3.20

path of the radiation is a power function of the temperature, it is natural to assume that as $u \to a_{1-}$ we have

$$K(u) = \tilde{a}_1 + K_0(a_1 - u)/(K_0 - 1), \quad 1 < K_0 = \text{const} < 2. \tag{3.6.5.2}$$

The thermal wave is said to be formed if it is represented by a continuous nonnegative solution $u(x, t, \varepsilon)$ to Eq. (3.6.5.1) such that for $a_0 \leqslant u(x, t, \varepsilon) \leqslant a_1$

$$\begin{gathered} u(x, t, \varepsilon) = a_0 + \delta(\varepsilon), \quad \delta(\varepsilon) > 0, \\ \delta(\varepsilon) = O(\varepsilon^m), \quad m > 0 \text{ for } x \in \Omega_1, \\ u(x, t, \varepsilon) \equiv a_1 \text{ for } x \in \Omega_2, \end{gathered} \tag{3.6.5.3}$$

where Ω_1 and Ω_2 are regions in R^3.

In the one-dimensional case, the solution constructed in this section is depicted in Figure 3.20. The problem of the propagation of a formed thermal wave in the one-dimensional case ($x \in R^1$) and with second-order time derivatives ignored has been considered in [3.35, 3.36], where the solution was constructed by matching the self-similar part describing the motion of the wavefront of the thermal wave with a certain constant describing the temperature in the inner region, and the graph of this solution resembles the one shown in Figure 3.20. On the whole, however, the constructed thermal wave does not satisfy the heat equation with constant coefficients discussed in [3.35, 3.36] due to the presence of discontinuous derivatives. The method of constructing the asymptotic solution described in the present paper is close to the one discussed in [3.3].

The main result of the present section can be formulated in the following

Theorem 3.6.5.1 *The asymptotic solution to the problem of propagation of a formed thermal wave is*

$$u(x,\ t,\ \varepsilon) = W_0(x,\ \tau) + \varepsilon W_1(x,\ \tau) + O(\varepsilon^2)|_{\tau=S(x,\ t)/\varepsilon} \tag{3.6.5.4}$$

with $W_0 = \chi(\tau)$, *where the function* $\chi(\tau)$ *is the solution to the boundary value problem for the ordinary differential equation*

$$\begin{gathered} b\frac{d\chi}{d\tau} + b^2\frac{d^2\chi}{d\tau^2} - \frac{d}{d\tau}\left(K(\chi)\frac{d\chi}{d\tau}\right) - F(\chi) = 0, \\ \chi|_{\tau\to-\infty} \to a_0, \quad \chi|_{\tau\to 0} \to a_1, \end{gathered} \tag{3.6.5.5}$$

and the following estimates hold true: $\chi = a_1 - O(\tau^\alpha)$ *as* $\tau \to 0$, $\alpha = 1/(1-q)$, $\alpha > 0$ *and* $\chi \sim a_0 + \exp(l_0\tau)$ *as* $\tau \to -\infty$ ($l_0 =$ const). *The function* $S(x,\ t)$ *has the form*

$$S(x,\ t) = b(t + \psi(x)) + \varphi_1(x)(t + \psi(x))^2,$$

where $b \geqslant 2\sqrt{dR/d\chi|_{\chi=a_0}}$ *and* $R = F(\chi)(K(\chi) - b^2)$, *and the function* $\psi(x)$ *satisfies the equation*

$$|\nabla\psi|^2 = (b^2\lambda(x,\ -\psi(x)))^{-1}. \tag{3.6.5.6}$$

The surface $\Gamma(t)$ *of weak discontinuity of the solution has the form*

$$\Gamma(t) = \{x\colon\ \psi(x) = -t\},$$

and the surface $\Gamma(0)$ *is considered fixed,*

$$\Omega_1 \subset \{x\colon\ \psi(x) < -t\},\ \Omega_2 \subset \{x\colon\ \psi(x) > -t\}.$$

The necessary condition for the existence of a type (3.6.4.4) *solution is* $b = \sqrt{K(a_1)}$.[11]

The function W_1 *has the form*

$$W_1 = -\frac{d\chi}{d\tau}\int_0^\tau \left[V\left(\frac{d\chi}{d\tau}\right)^{-2}\int_{-\infty}^{\tau'}\frac{f\,(d\chi/d\xi)\,d\xi}{V(K(\chi)-b^2)}\right]d\tau', \tag{3.6.5.7}$$

where

$$\begin{aligned} f = {} & 2\frac{\varphi_1(x)}{b}\tau\frac{d\chi}{d\tau} + 4\varphi_1(x)\tau\frac{d^2\chi}{d\tau^2} + 2\varphi_1(x)\frac{d\chi}{d\tau} \\ & - b\langle\nabla\psi,\ \nabla\lambda|_{t=-\psi(x)}\rangle K(\chi)\frac{d\chi}{d\tau} \\ & - K(\chi)\frac{d\chi}{d\tau}\lambda(x,\ -\psi(x))\,[b\nabla^2\psi + 2\varphi_1|\nabla\psi|^2 + 2\langle\nabla\varphi,\ \nabla\psi\rangle] \end{aligned}$$

---

[11] This condition means that the speed of the thermal wave is equal to the effective speed of sound inside the wave and was first proposed on physical grounds by A.S. Kompaneyets.

$$-\frac{\partial\lambda}{\partial t}\Big|_{t=-\psi(x)}\tau\frac{\partial}{\partial\tau}\left(K(\chi)\frac{d\chi}{d\tau}\right)b\,|\nabla\psi|^2$$
$$-\lambda(x,\ -\psi(x))\frac{\partial}{\partial\tau}\left(K(\chi)\frac{d\chi}{d\tau}\right)2\tau\left(\langle\nabla\psi,\ \nabla\varphi\rangle\right.$$
$$\left.+\frac{2\varphi_1}{b}|\nabla\psi|^2\right), \tag{3.6.5.8}$$

and V, *the Wronskian of the equation for finding* W_1, *has the form*

$$V=(K(\chi)-b^2)^{-2}\exp\left(b\int\frac{d\tau}{(K(\chi)-b^2)}\right). \tag{3.6.5.9}$$

The boundary conditions for W_1 *have the form*

$$W_1|_{\tau\to-\infty}=0,\quad W_1|_{\tau=0}=0.$$

The equation for finding $\varphi_1(x)$ *follows from the necessary condition of solvability, which for* $q<1$ *and* $k+q>2$ *has the form*

$$\lim_{\tau\to0}\left[\int_\tau^0 V\left(\frac{d\chi}{d\tau'}\right)^2\left(\int_{-\infty}^{\tau'}\frac{(d\chi/d\tau)\,f\,d\tau}{V(K(\chi)-b^2)}\right)d\tau'\right]=0.$$

The function f *is given by formula* (3.6.5.8).

If we put

$$I_0=\int_\tau^0 M_1(\tau')\left(\int_{-\infty}^{\tau'}\frac{\tau(d\chi/d\tau)^2\,d\tau}{M(\tau)}\right)d\tau',$$
$$I_1=\int_\tau^0 M_1(\tau')\left(\int_{-\infty}^{\tau'}\frac{(d\chi/d\tau)^2\,d\tau}{M(\tau)}\right)d\tau',$$
$$I_2=\int_\tau^0 M_1(\tau')\left(\int_{-\infty}^{\tau'}\frac{K(\chi)(d\chi/d\tau)^2\,d\tau}{M(\tau)}\right)d\tau', \tag{3.6.5.10}$$
$$I_3=\int_\tau^0 M_1(\tau')\left(\int_{-\infty}^{\tau'}\frac{\tau\frac{d}{d\tau}\left(K(\chi)\frac{d\chi}{d\tau}\right)\frac{d\chi}{d\tau}\,d\tau}{M(\tau)}\right)d\tau',$$
$$I_4=\int_\tau^0 M_1(\tau')\left(\int_{-\infty}^{\tau'}\frac{\tau(d\chi/d\tau)(d^2\chi/d\tau^2)\,d\tau}{M(\tau)}\right)d\tau',$$
$$M_1(\tau)=V\left(\frac{d\chi}{d\tau}\right)^{-2},\quad M(\tau)=V(K(\chi)-b^2),$$

then

$$\varphi_1=\lim_{\tau\to0}\left[\left\{\frac{2}{b}I_0+4I_4+2I_1-I_2\lambda(x,\ -\psi(x))\,2\,|\nabla\psi|^2\right.\right.$$
$$\left.-\lambda(x,\ -\psi(x))\frac{4}{b}|\nabla\psi|^2 I_3\right\}^{-1}$$

$$\times \Big\{ b \langle \nabla\psi, \nabla\lambda|_{t=-\psi(x)}\rangle I_2 + \lambda(x, -\psi(x))(b\nabla^2\psi + 2\langle\nabla\varphi, \nabla\psi\rangle) I_2 + I_3 b\, |\nabla\psi|^2 \frac{\partial\lambda}{\partial t}\Big|_{t=-\psi(x)} + I_3\lambda(x, -\psi(x))\, 2\langle\nabla\varphi, \nabla\psi\rangle\Big\}\Big]. \tag{3.6.5.11}$$

Proof. The main equation obtained as a result of substituting solution (3.6.5.4) into Eq. (3.6.5.1) has the form

$$\left(\frac{\partial W_0}{\partial\tau} + \varepsilon\frac{\partial W_1}{\partial\tau}\right)\frac{\partial S}{\partial\tau} + \left(\frac{\partial^2 W_0}{\partial\tau^2} + \varepsilon\frac{\partial^2 W_1}{\partial\tau^2}\right)\left(\frac{\partial S}{\partial t}\right)^2 + \varepsilon\frac{\partial W_0}{\partial t} + \varepsilon\frac{\partial W_0}{\partial\tau}\frac{\partial^2 S}{\partial t^2} + \ldots + O(\varepsilon^2)$$
$$- \varepsilon\langle\nabla\lambda(x, t), \nabla S\rangle K(W_0)\frac{\partial W_0}{\partial\tau} - \varepsilon\lambda(x, t) K(W_0)\frac{\partial W_0}{\partial\tau}\nabla^2 S$$
$$- \lambda(x, t)\frac{\partial}{\partial\tau}\left\{K(W_0)\frac{\partial W_0}{\partial\tau} + \varepsilon\frac{\partial}{\partial\tau}(K(W_0)W_1) + \ldots + O(\varepsilon^2)\right\} - F(W_0) - \varepsilon\frac{dF}{dW_0}W_1 = 0. \tag{3.6.5.12}$$

The function $S(x, t, \varepsilon)$ has the form

$$S(x, t, \varepsilon) = \varphi(x)(t + \psi(x)) + \varphi_1(x)\,(t + \psi(x))^2$$

(see [3.3]), with the following estimates holding true:

$$\frac{\partial S}{\partial t} = \left[\varphi(x) + 2\varphi_1\frac{\varepsilon\tau}{\varphi(x)} + O((\varepsilon\tau)^2)\right]_{\tau=S/\varepsilon},$$
$$\nabla S = \left[\varphi(x)\nabla\psi + \frac{\varepsilon\tau}{\varphi}\left(\nabla\varphi + \frac{2\varphi_1}{\varphi}\nabla\psi\right) + O((\varepsilon\tau)^2)\right]\Big|_{\tau=S/\varepsilon},$$
$$(\nabla S)^2 = \Big[\varphi^2|\nabla\psi|^2 + 2\varepsilon\tau\langle\nabla\psi, \nabla\varphi\rangle + 4\varepsilon\tau\frac{\varphi_1}{\varphi}|\nabla\psi|^2 + O((\varepsilon\tau)^2)\Big]\Big|_{\tau=S/\varepsilon},$$
$$\nabla^2 S = \varphi\nabla^2\psi + 2\langle\nabla\varphi, \nabla\psi\rangle + 2\varphi_1(\nabla\psi)^2 + O(\varepsilon\tau)|_{\tau=S/\varepsilon},$$
$$\frac{\partial^2 S}{\partial t^2} = 2\varphi_1 + O(\tau\varepsilon)|_{\tau=S/\varepsilon}.$$

Nullifying the sum of coefficients of ε^0 and of ε^1 in (3.6.5.12), we arrive at the two following equations

$$\frac{\partial W_0}{\partial\tau}\varphi(x) + \frac{\partial^2 W_0}{\partial\tau^2}\varphi^2(x) - \lambda(x, -\psi(x))\frac{\partial}{\partial\tau}\left(K(W_0)\frac{\partial W_0}{\partial\tau}\right)\varphi^2(\nabla\psi)^2 - F(W_0) = 0, \tag{3.6.5.13}$$

$$\varphi\frac{\partial W_1}{\partial\tau} + \varphi^2\frac{\partial^2 W_1}{\partial\tau^2} - \lambda(x, -\psi(x))\frac{\partial^2}{\partial\tau^2}(K(W_0)W_1)\varphi^2|\nabla\psi|^2 - \frac{dF}{dW_0}W_1 = -f, \tag{3.6.5.14}$$

where

$$f=\frac{2\varphi_1}{\varphi}\tau\frac{\partial W_0}{\partial\tau}+4\varphi_1\tau\frac{\partial^2 W_0}{\partial\tau^2}+2\varphi_1\frac{\partial W_0}{\partial\tau}$$
$$-\varphi\langle\nabla\psi,\ \nabla\lambda|_{t=-\psi(x)}\rangle K(W_0)\frac{\partial W_0}{\partial\tau}$$
$$-K(W_0)\frac{\partial W_0}{\partial\tau}\lambda(x,\ -\psi(x))[\varphi\nabla^2\psi$$
$$+2\langle\nabla\varphi,\ \nabla\psi\rangle+2\varphi_1(\nabla\psi)^2]$$
$$-\frac{\partial}{\partial\tau}\left(K(W_0)\frac{\partial W_0}{\partial\tau}\right)\tau\left[\frac{\partial\lambda}{\partial t}\Big|_{t=-\psi}\varphi^2(\nabla\psi)^2\right.$$
$$\left.+\lambda(x,\ -\psi(x))\,2\left(\langle\nabla\varphi,\ \nabla\psi\rangle+\frac{2\varphi_1}{\varphi}|\nabla\psi|^2\right)\right].$$

Assuming that $b=\varphi(x)$ and $\lambda(x, -\psi(x))\,b^2(\nabla\psi)^2=1$, we find that the function $W_0=\chi(\tau)$ is the solution to problem (3.6.5.5). The change of variables carried out in Section 3.2 (see Eq. (3.2.2.3)) results in the following equation:

$$b\frac{d\Theta}{d\xi}-\frac{d^2\Theta}{d\xi^2}-R(\Theta)=0,$$

where

$$R(\Theta)=(K(a_1-\Theta)-b^2)\widetilde{F}(\Theta),\quad b>0,$$
$$\Theta|_{\xi\to-\infty}=a_1,\quad \Theta|_{\xi\to+\infty}=a_0,\quad d\Theta/d\xi\geqslant 0.$$

This equation was used in Section 3.2 in connection with the case where

$$dR(\Theta)/d\Theta\,|_{\Theta=0}=0$$

(see also [3.3]).

Introduction of $V=a_1-\Theta$ yields the following equation:

$$b\frac{dV}{d\xi}-\frac{d^2V}{d\xi^2}+\widetilde{R}(V)=0,$$
$$V|_{\xi\to-\infty}=a_1-a_0,\quad V|_{\xi\to+\infty}=0,\quad dV/d\xi\leqslant 0,$$

where the function V possesses the following estimates:

$$V\sim 1/\xi \text{ as } \xi\to\infty$$
$$V\sim a_1-a_0-\exp(l_0\xi) \text{ as } \xi\to-\infty,$$

where

$$l_0=b/2-\sqrt{b^2/4-dR/d\Theta|_{\Theta=a_0}}.$$

Function ψ constitutes a solution to the problem

$$|\nabla\psi|^2=[b^2\lambda(x,\ -\psi(x))]^{-1}.$$

At time t the wavefront is described by the reference[12]

$$\Gamma(t) = \{x\colon \psi(x) = -t\},$$

which separates the regions

$$\Omega_1 \subset \{x\colon \psi(x) > -t\},\quad \Omega_2 \subset \{x\colon \psi(x) \leqslant -t\}.$$

Employing (3.2.2.3), we can transform these estimates into

$$\chi = a_1 - O(\tau^{\alpha}) \text{ as } \tau \to 0 \quad (\alpha = 1/(1-q)),$$
$$\chi \sim a_0 + \exp(l_0\tau) \text{ as } \tau \to -\infty.$$

Thus, there exists a one-sided localized solution to problem (3.6.5.5) related to a traveling wave. Let us rewrite the equation for finding W_1 in the form

$$\frac{\partial^2 W_1}{\partial\tau^2} - \frac{\partial W_1}{\partial\tau}\,\frac{b - 2\partial K(W_0)/\partial\tau}{K(W_0) - b^2} + W_1\,\frac{\partial^2 K(W_0)/\partial\tau^2 + dF(W_0)/dW_0}{K(W_0) - b^2}$$
$$= -\frac{f}{K(W_0) - b^2}, \tag{3.6.5.15}$$
$$W_1|_{\tau\to+\infty} = 0, \quad W_1|_{\tau\to-\infty} = 0,$$

where f is defined in (3.6.5.14).

The Wronskian of the ordinary differential equation in variable τ has the form (3.6.5.9). Equation (3.6.5.15) can be analyzed in the same manner as in [3.3]. The solution has the form (3.6.5.7).

The function φ_1 is found from the necessary condition of solvability of the problem and has the form (3.6.5.11). The integrals I_i, $i = 1, 2, 3, 4$ are found by formulas (3.6.5.10).

3.6.6 Propagation of Nonlinear Thermal Waves (Continued)

The algorithm for constructing an asymptotic solution changes little when the thermal conductivity coefficient increases with u. This modified algorithm is given below.

Let us consider problem (3.6.5.1) for the case where $K(u)$ has the form

$$K(u) = \tilde{a}_1 - K_0(a_1 - u)^{k-1}, \tag{3.6.6.1}$$

where a_1 and K_0 are positive constants. We will seek the asymptotic solution to problem (3.6.5.1), (3.6.6.1) in the form

$$u(x, t, \varepsilon) = W_0(x, \tau) + \varepsilon W_1(x, \tau) + O(\varepsilon^2)|_{\tau = S(x, t, \varepsilon)/\varepsilon} \tag{3.6.6.1$'$}$$

The main assertion of this section is formulated in the following

---

[12] Here and in what follows it is assumed that on a surface $\Gamma(0)$ the trajectories of the corresponding system of ordinary differential equations are projected in a unique manner on R^2_x.

Theorem 3.6.6.1 *Problem* (3.6.5.1), (3.6.6.1) *has an asymptotic solution of the form* (3.6.6.1′) *with* $W_0 \equiv \chi(\tau)$, *where the function* $\chi(\tau)$ *is the solution to the boundary value problem*

$$b\frac{d\chi}{d\tau}+b^2\frac{d^2\chi}{d\tau^2}-\frac{d}{d\tau}\left(K(\chi)\frac{d\chi}{d\tau}\right)-F(\chi)=0, \qquad (3.6.6.2)$$

$$\chi|_{\tau\to-\infty}=a_0, \quad \chi|_{\tau\to\infty}=a_1,$$

where

$$b\geqslant 2\sqrt{|dR/d\Theta|_{\Theta=a_1}|}, \quad R=F(\chi)(K(\chi)-b^2),$$

and the following estimates hold true:

$$\chi\sim a_1-O(\tau^{\alpha}) \text{ as } \tau\to 0, \ \alpha=1/(k-1), \ \alpha>0,$$
$$\chi\sim a_0+\exp(l_0\tau) \text{ as } \tau\to-\infty. \qquad (3.6.6.3)$$

The function $S(x, t)$ *has the form*

$$S(x, t) = \varphi(x)(x+\psi(x)) + \varphi_1(x)(x+\psi(x))^2,$$

with $\psi(x)$ *satisfying the equation*

$$|\nabla\psi|^2 = (b^2\lambda(x, -\psi(x)))^{-1}.$$

The surface $\Gamma(t)$ *of weak (removable) discontinuity of the solution has the form*

$$\Gamma(t)=\{x:\ \psi(x)=-t\},$$

with $\Gamma(0)$ *assumed given,*

$$\Omega_1\subset\{x:\ \psi(x)<-t\}, \quad \Omega_2\subset\{x:\ \psi(x)>-t\}.$$

The necessary condition for the existence of a solution of the (3.6.5.4) *type is* $b=\sqrt{K(a_1)}$.

The function W_1 *has the form* (3.6.5.7). *The function* f *is defined in* (3.6.5.8), *and the necessary condition for the solvability of the problem in the case* (3.6.6.1) *has the form*

$$\int_{-\infty}^{0}\frac{f(d\chi/d\tau)\,d\tau}{\sqrt{(K(\chi)-b^2)}}=0.$$

Function φ_1 *is defined in* (3.6.5.11) *with*

$$I_0=\int_{-\infty}^{0}\frac{\tau(d\chi/d\tau)^2\,d\tau}{M}, \quad I_1=\int_{-\infty}^{0}\frac{(d\chi/d\tau)^2\,d\tau}{M},$$

$$I_2=\int_{-\infty}^{0}\frac{K(\chi)(d\chi/d\tau)^2\,d\tau}{M}, \qquad (3.6.6.4)$$

$$I_3 = \int_{-\infty}^{0} \frac{\tau \frac{d}{d\tau}\left(K(\chi)\frac{d\chi}{d\tau}\right)\frac{d\chi}{d\tau}\,d\tau}{M},$$

$$I_4 = \int_{-\infty}^{0} \frac{\tau\,(d\chi/d\tau)\,(d^2\chi/d\tau^2)\,d\tau}{M}, \quad M = V\,(K(\chi) - b^2).$$

Proof. Equations (3.6.5.12)-(3.6.5.14) remain valid, $W_0 \equiv \chi(\tau)$. By substituting $a_1 - \Theta$ for χ and applying (3.2.2.3) we can transform Eq. (3.6.6.2) into

$$b\frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} - R(\Theta) = 0, \quad R(\Theta) = (K(\Theta) - b^2)\,F(\Theta),$$

$$\Theta|_{\xi\to-\infty} = a_1, \quad \Theta|_{\xi\to+\infty} = a_0, \quad d\Theta/d\xi < 0.$$

This is the Zeldovich equation studied in Section 3.2 (see Theorem 3.2.2.2). Indeed, the substitution $V = a_1 - \Theta$ yields

$$b\frac{dV}{d\xi} - \frac{d^2V}{d\xi^2} + \tilde{R}(V) = 0,$$

where $\tilde{R}(V) < 0$ and $V \in [0, a_1 - a_0]$, with

$$V|_{\xi\to+\infty} = a_1 - a_0, \quad V|_{\xi\to-\infty} = 0, \quad dV/d\xi > 0.$$

The function V has the following estimates:

$$V = a_1 - a_0 - \exp(-l_1\xi) \text{ as } \xi \to \infty,$$

where

$$l_1 = -b/2 + \sqrt{b^2/4 + |dR/dV|_{V=a_1-a_0}|},$$

$$b > 2\sqrt{|dR/dV|_{V=0}|},$$

and

$$V = \exp(l\xi) \text{ as } \xi \to -\infty,$$

where

$$l = b/2 - \sqrt{b^2/4 - |dR/dV|_{V=0}|}.$$

From this follow the estimates for $\chi(\tau)$:

$$\chi(\tau) \sim a_1 - O(\tau^\alpha) \quad \text{as } \tau \to 0, \ \alpha = 1/(k-1),$$

$$\chi(\tau) \sim a_0 + \exp(l_0\tau) \quad \text{as } \tau \to -\infty \ (l_0 = \text{const}).$$

The remainder of the proof is the same as the proof of Theorem 3.6.5.1. The necessary solvability condition for $q \geqslant 1$ or $q < 1$ and $K(u) < b^2$ has the form

$$\int_{-\infty}^{0} \frac{f\,(d\chi/d\tau)\,d\tau}{V\,(K(\chi) - b^2)} = 0.$$

The function φ_1 is defined in (3.6.5.1) in which the I_i are calculated via (3.6.6.4). As shown in Section 3.2, all these integrals exist. The proof of the theorem is complete.

Let us give some examples of the solution of model problems involving Eq. (3.6.5.5). This equation is similar to the one that de-

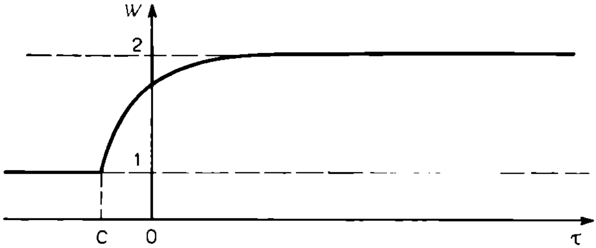


Fig. 3.21

scribes diffusion of light in an active medium and the heat equation, for one thing, the heat transfer in a superconducting matrix (see (3.1.2.2) and (3.1.2.1)).

Suppose that

$$K(u) = 1 + \sqrt{u-1}, \quad F(u) = 2\sqrt{u-1}\,(4\sqrt{u-1} - 3u + 2); \qquad (3.6.6.5)$$

In this case $a_0 = 1$ and $a_1 = 2$ are the roots of the equation $F(u) = 0$, $q = 1/2$, $k = 3/2$, that is,

$$K(u) \sim (u-1)^{k-1} \quad 1, \quad u \to 1; \quad F(u) \sim (u-1)^q, \quad u \to 1 \neq 0. \qquad (3.6.6.6)$$

The equation

$$\varepsilon \frac{\partial u}{\partial t} + \varepsilon^2 \frac{\partial^2 u}{\partial t^2} - \varepsilon^2 \frac{\partial}{\partial x}\left(K(u)\frac{\partial u}{\partial x}\right) - F(u) = 0 \qquad (3.6.6.7)$$

has an exact solution of the form $u(x, t, \varepsilon) = W(\tau)$, with $\tau = x - bt$. The function $W(\tau)$ satisfies the standard equation

$$b\frac{dW}{d\tau} - \frac{d}{d\tau}\left((K(W) - b^2)\frac{dW}{d\tau}\right) - F(W) = 0, \qquad (3.6.6.8)$$

whose properties were studied in Section 3.2.

At $b = 1$ the solution to Eq. (3.6.5.5) has the form

$$W(\tau) = \begin{cases} 1 + (1 - \exp(-\tau - C))^2 & \text{if } \tau > -1 \ (C = \text{const} > 0), \\ 1 & \text{if } \tau < -1. \end{cases}$$

The graph representing this solution is shown in Figure 3.21. Suppose that in Eq. (3.6.6.8)

$$K(u)=4+\sqrt{2-u},\quad F(u)=2\sqrt{2-u}\,(9-6\sqrt{2-u}-3u), \qquad (3.6.6.9)$$

where $a_0 = 1$, $a_1 = 2$, $q = 1/2$, and $k = 3/2$. Then the solution to Eq. (3.6.6.8) has the form

$$W=\begin{cases} 2 & \text{if} \quad \tau>0 \\ 2-(1-\exp\tau)^2 & \text{if} \quad \tau<0. \end{cases} \qquad (3.6.6.10)$$

The graph of this solution is shown in Figure 3.20.

3.7 Solution of Equations of the Ginzburg-Landau Type. Waves in Ferromagnetic Substances

In this chapter we build the asymptotic solution to parabolic semilinear equations of the Ginzburg-Landau type, a solution that describes the spin waves in a ferromagnetic substance, one of the initial stages in the development of turbulence on the surface of a liquid under parametric excitation, and some other phenomena.

The nonlinear problem of regular structures emerging on the surface of a liquid and of the complication of these structures and their stochastization is important to many fields of physics, to such phenomena as the Langmuir waves in plasmas, the Tolmien-Schlichting waves in fluid mechanics, waves on shallow water, spin waves in ferromagnetic substances, and waves on the surface of liquid insulators placed in an electric field or on the surface of ferromagnetic liquids placed in a variable magnetic field. Similar phenomena are observed when surfaces are exposed to ion and laser beams [3.23].

A theory concerning these phenomena and based on the approach used in studying capillary waves has been developed in [3.21, 3.22]. The emergence of structures on the surface of a liquid, the so-called Faraday ripples, is due to the parametric generation of counter-running waves. The conditions under which the complex-valued amplitude of such a wave varies smoothly in space have been formulated in [3.23]. The reader interested in the theory of such waves can also refer to [3.37-3.39].

When a pair of such waves interact, the graph of the absolute value of the complex-valued amplitude shows, in a certain interval of values of the parameters (see [3.23]), a number of sufficiently stable grooves, or bands (see Figure 3.7), while as the system evolves a sequence of dark bands emerges, and against the background of these dark bands smaller structures develop. Apparently, these structures can be described in the nonlinear approximation because the model describing this system has a small dimensionless parameter.

The system of equations for the envelopes has the form

$$\frac{\partial a_\pm}{\partial t} \pm v_g \frac{\partial a_\pm}{\partial x} - \frac{i}{4}\frac{v_g}{k}\frac{\partial^2 a_\pm}{\partial x^2} - i\frac{v_g}{2k}\frac{\partial^2 a_\pm}{\partial y^2} + \gamma a_\pm$$
$$= i(H + Fb_+b_-)a^*_\mp + ia_\pm [T|a_\pm|^2 + S|a_\mp|^2 + R(|b_+|^2 + |b_-|^2)], \tag{3.7.0.1}$$
$$\frac{\partial b_\pm}{\partial t} \pm v_g \frac{\partial b_\pm}{\partial y} - \frac{i}{4}\frac{v_g}{k}\frac{\partial^2 b_\pm}{\partial y^2} - i\frac{v_g}{2k}\frac{\partial^2 b_\pm}{\partial x^2} + \gamma b_\pm$$
$$= i(H + Fa_+a_-)b^*_\mp + ib_\pm [T|b_\pm|^2 + S|b_\mp|^2 + R(|a_+|^2 + |a_-|^2)],$$

where γ is the damping constant, a and b the dimensional complex-valued amplitudes of the wave, v_g the group velocity of the wave, k the wave number, and F, H, R, S, and T are constants (see Section 3.1).

The characteristic values of the parameters for steady-state capillary waves are $F = 0.33\,\omega k^2$, $H = [k/(4\omega)]\,c$, $R = -0.18\omega k^2$, $S = 0.625\omega k^2$, $T = 0.0625\omega k^2$, $\gamma = 2\nu k^2$, $v_g/\lambda\gamma = 3.4$, and $k = 20 \times 10^{-2}\ \mathrm{m}^{-1}$, with λ the wavelength and ω the frequency.

The excitation of a pair of waves modulated across the wavefront (i.e. along the y-axis) is described by an analog of the Ginzburg-Landau equation [3.23], which in dimensionless variables has the form

$$\frac{\partial u}{\partial \tau} - i\sigma\varepsilon^2 \frac{\partial^2 u}{\partial \eta^2} + u - ihu^* - iu|u| + i\sigma u = 0, \tag{3.7.0.2}$$

where σ is a constant equal to β/γ (see Section 3.1), $\eta = t/(\lambda m)$, $\tau = t\gamma$, $u = a[(S+T)/\gamma]^{1/2}$, $h = H/\gamma = 1 + h_0$, $h_0 \sim \varepsilon$, $l = v_g(2k\lambda^2\gamma m^2)^{-1}$, and m an integer. Here $\varepsilon = \sqrt{l\sigma} = [v_g(2k\lambda^2 \times \gamma\sigma m^2)^{-1}]^{1/2}$ is a small parameter ($\varepsilon < 1$).

The boundary conditions are

$$u|_{\eta\to+\infty} = 1, \quad u|_{\eta\to-\infty} = 1. \tag{3.7.0.2'}$$

The following theorem holds true:

Theorem 3.7.0.1 *Problem* (3.7.0.2), (3.7.0.2′) *has an asymptotic solution of the form*

$$u = \sqrt{\sigma/2}\,(w_2 + iw_2),$$
$$w_2 = \sum_{l=1}^{n} \omega_-\left(\frac{\eta+\eta_l}{\varepsilon}\sqrt{\sigma}\right)$$

where $\eta_l = l\delta_2$, $\delta_2 > 2\delta$, *with* δ *the period of the structure,*

$$\omega_-(\xi) = -\frac{\sqrt{2}}{\cosh \xi}, \quad \xi = \frac{\eta\sqrt{\sigma}}{\varepsilon},$$

and the following estimate holds true:

$$-\sqrt{2}N\varepsilon\ln\varepsilon < \delta_2 < \frac{1}{m\,(2\varepsilon^{N_1})^{1/4}}\left(\frac{\pi v_g}{\lambda\gamma}\right)^{1/2}, \quad N_1, \quad N>1.$$

Proof. Equation (3.7.0.2) implies the equation (see Section 3.1.3)

$$-\frac{\partial^2\omega}{\partial\xi^2}+\omega(1-\omega^2)=0, \tag{3.7.0.3}$$

where $u = n(1+i) = (1+i)\sqrt{\sigma/2}\,\omega(\xi)$. Equation (3.7.0.3) has two trivial stable solutions $\omega = \pm 1$ and the solution

$$\omega_\pm = \pm\sqrt{2}/\cosh\xi. \tag{3.7.0.4}$$

The graph of function (3.7.0.4) is given in Figure 3.9 (curve *1*). A constructive method of building solutions of the (3.7.0.4) type and time-dependent solutions for more general quasilinear parabolic equations is discussed in [3.3]. The function (3.7.0.4) is exponentially close to 0 at a fairly small distance from point $\eta = 0$; precisely, for ω_- the following estimates hold true:

$$\omega_- = \begin{cases} 1-\varepsilon^N+o(\varepsilon^N), & N>1, \quad \xi>\delta\sqrt{\sigma}/\varepsilon = -N\ln\varepsilon, \\ -1+\varepsilon^N+o(\varepsilon^N), & N>1, \quad \xi<N\ln\varepsilon. \end{cases} \tag{3.7.0.5}$$

Similar estimates exist in relation to solution ω_+.

Equation (3.7.0.3) is invariant with respect to the translation group; hence, the function

$$\begin{aligned} \omega_+(\eta/\varepsilon) &= \frac{1-\exp((\eta+\eta_1)\sqrt{2}/\varepsilon)}{1+\exp((\eta+\eta_1)\sqrt{2}/\varepsilon)}, \\ \omega_- &= -\sqrt{2}/\cosh(\xi+\xi_1), \quad \xi_1=\text{const} \end{aligned} \tag{3.7.0.6}$$

is also a solution to this equation. The graph of function $\tilde{\omega}_-$ is shown in Figure 3.9 (curve *2*). Obviously, it is easy to select constants N, δ, and δ_1 in such a manner that all the necessary conditions are satisfied. The function $u(\eta/\varepsilon) + u((\eta+\eta_l)/\varepsilon)$ is also, to within $O(\varepsilon^N)$, an asymptotic solution to (3.7.0.2). The graph of this function is shown in Figure 3.8 (the cross section x = const). Apparently, the sequence of grooves can be described by the formula

$$u = (1+i)\sum_{l=1}^{n}\omega_-(\xi+\xi_l) \tag{3.7.0.7}$$

where $\xi_l = l\xi_1$, $\xi_1 = \delta/\varepsilon$, $\delta_2 < 2\delta$, and δ_2 is the period of the structure.

The steady-state distribution of the amplitude, that is, the graph of the functions $|u| = \sigma|W_j|$, $j = 1, 2$, is shown in Figure 3.7. The functions

$$u = (\sqrt{\sigma}/2)(W_j + iW_j) \tag{3.7.0.8}$$

are asymptotic solutions to Eq. (3.7.0.2) to within $O(\varepsilon^N)$ uniformly in variable η. The solutions thus constructed are stable within a certain interval of parameters. Their stability has been studied in [3.23], where the solution to Eq. (3.7.0.2) was sought in the form $u = u_{st.-st} + \psi(\eta, \tau)$ and a linearized equation for ψ was studied. The solution of this linearized equation was sought in the form of plane waves. The characteristic modulation period in variable η must not exceed $\delta_2 = [(2h_0)^{-1/4}/m] (\pi v_g/(\gamma\lambda))^{1/2}$. Thus, the value of the period has the following upper and lower bounds:

$$-\sqrt{2}N\varepsilon \ln \varepsilon < \delta_2 < [(2\varepsilon^{N_1})^{-1/4}/m] (\pi v_g/(\gamma\lambda))^{1/2}.$$

In conclusion we give the following

Theorem 3.7.0.2 *The semi-linear equation*

$$\varepsilon \frac{\partial u}{\partial t} - \varepsilon^2 \frac{\partial^2 u}{\partial x^2} - u(1-u^2) = 0, \quad u \in C^\infty \tag{3.7.0.8'}$$

(ε is a small parameter), has an exact two-phase solution of the form

$$u(x, t, \varepsilon) = \frac{1 - A\exp(\pm\sqrt{2}x/\varepsilon)}{1 + A\exp(\pm\sqrt{2}x/\varepsilon) + B\exp\{(\pm x/\sqrt{2} - 3t/2)/\varepsilon\}} \tag{3.7.0.8''}$$

with A and B constants.

Proof. The proof of this theorem is given in [3.3] and can be carried out by directly substituting into Eq. (3.7.0.8′) a function of the form $u(x, t, \varepsilon) = F(x, t, \varepsilon)/G(x, t, \varepsilon)$.

Function (3.7.0.8″) depends on two "phases" (functions), $\pm\sqrt{2}x/\varepsilon$ and $(\pm x\sqrt{2} - 3t/2)/\varepsilon$, and satisfies, say, for the plus sign in the exponents, the following conditions:

$$u|_{x\to\infty} \to -1, \quad u|_{x\to-\infty} \to 1.$$

Two-phase solutions to semi-linear parabolic equations have, apparently, not been studied up till now.

3.8 Asymptotic and Characteristic Exact Solutions to Semi-Linear and Quasilinear Parabolic and Hyperbolic Equations

(Wave Type Solutions; Synergets Bounded as $\varepsilon \to 0$)

A number of papers and books by the authors of the present article (primarily [3.3]) and other investigators, say [3.26-3.30, 3.37-3.40], devoted to the study of properties of semi-linear and quasilinear parabolic and hyperbolic equations involving a small parameter ε have made it possible to clearly specify some classes of solutions to such equations. Here we give the solutions that are characteristic only of two classes.

In this section we deal with nonlocalized synergets of semi-linear and quasi-linear hyperbolic and parabolic equations that are bounded as $\varepsilon \to 0$. For comparison, in Subsections 1.6, 1.7, and 1.8 we give without proof the formulas for localized solutions to singular parabolic equations bounded as $\varepsilon \to 0$ [3.3]. None but the one-dimensional case will be considered.

1.1. *Semi-linear parabolic equations with constant coefficients:*

$$\frac{\partial z}{\partial t}-\frac{\partial^2 z}{\partial t^2}-\mathcal{F}(z)=0, \quad \mathcal{F}(z)\in C^1\ [a_1,\ a_2]. \qquad (3.8.0.1)$$

The equation $\mathcal{F}(z)=0$ has two roots: $z=a_i$, $i=1,\ 2$. (The scaling transformation $z=u(a_2-a_1)+a_1$ reduces this equation to $F(u)=0$ with roots $u=0$ and $u=1$.) The characteristic exact solutions have the form $u(x,\ t)=\Theta(\xi)$, $\xi=\alpha x+bt$. Depending on the properties of $F(u)$, Eqs. (3.8.0.1) are defined as follows:

1.1.A. Kolmogorov-Petrovskii-Piskunov equations (KPP equations):

$$F(0)=0,\ \ F(1)=0,\ \ \left.\frac{dF}{du}\right|_{u=0}>0,\ \ \left.\frac{dF}{du}\right|_{u=1}<0. \qquad (3.8.0.2)$$

A.1. $F=u(1-u^\nu)$, $\Theta=[e^\xi/(1+e^\xi)]^{2/\nu}$, $\alpha=\nu/\sqrt{4+2\nu}$, $b=(4+\nu)\nu/(4+2\nu)$ (see also p. 184 in [3.3]).

A.2. $F=(u-1)(1-(1-u)^\nu)$, $\Theta=[(1+e^\xi)^{2/\nu}-e^{2\xi/\nu}]/(1+e^\xi)^{2/\nu}$, $\alpha=\nu/\sqrt{4+2\nu}$, $b=-\nu(4+\nu)/(4+2\nu)$ (see also p. 184 in [3.3]).

A.3. $F=u(1-u^\nu)$, $\Theta=\omega^\delta(\xi)$, $\alpha=1$, $\delta>0$, $\nu>0$, $A_0\xi=\int\limits_0^\omega[\omega(1-\omega^{\delta\nu/2})]^{-1}d\omega$, $A_0=[-(2+\delta\nu)\pm(16-12\delta+\delta^2\nu+\delta^3\nu^2)^{1/2}]\times[4(\delta-1)]^{-1}$, $b=-\delta\nu/2$ (see [3.37]).

1.1.B. Zeldovich equations:

$$F(0)=0,\ \ F(1)=0,\ \ \left.\frac{dF}{du}\right|_{u=0}=0,\ \ \left.\frac{dF}{du}\right|_{u=1}<0. \qquad (3.8.0.3)$$

B.1. $F(u)=u^2(1-u)$, $\Theta=[1-e^{-\xi/\sqrt{2}}]^{-1}$, $\alpha=\pm 1$, $b=1/\sqrt{2}$ (see p. 198 in [3.3]).

1.1.C. Semyonov equations (Fitzhugh-Nagumo equations):

$$\begin{gathered}F(0)=0,\ \ F(a_1)=0,\ \ F(1)=0,\ \ a_1\in(0,\ 1),\\ \left.\frac{dF}{du}\right|_{u=0}<0,\ \ \left.\frac{dF}{du}\right|_{u=a_1}>0,\ \ \left.\frac{dF}{du}\right|_{u=1}<0.\end{gathered} \qquad (3.8.0.4)$$

C.1. Two wave-type solutions: $F=-\mu u^3+(\mu+\nu)u^2-\nu u$, $a_1=\nu/\mu<1$, $u=(1+e^\xi)^{-1}$, $\alpha=\pm\sqrt{\mu/2}$, $b=\nu-\mu/2$ (see p. 193 in [3.3]); $u=a_1(1+e^\xi)^{-1}$, $\alpha=\pm\nu(2\mu)^{-1/2}$, $b=\nu(2\mu-\nu)\times(2\mu)^{-1}$ (see p. 197 in [3.3]).

1.1.D. If condition $F\in C^1[0,\ 1]$ is not met, Eqs. (3.8.0.1) may have localized solutions.

D.1. If $F(0)=0$, $F(1)=0$, $\left.\frac{dF}{du}\right|_{u=0}=-\infty$, $\left.\frac{dF}{du}\right|_{u=1}>0$, say, $F=-\gamma u^{q}(1-u^{(1-q)/2})$, $\gamma<0$, $0<q<1$, then

$$\Theta=\begin{cases}\left(1-\exp\left[-\frac{1-q}{1+q}\gamma\xi/b\right]\right)^{2/(1-q)} & \text{if } \tau<0,\\ 0 \quad \text{if} \quad \tau>0,\end{cases}$$

$$b=[-2(1+q)^{-1}\gamma]^{1/2},\ \alpha=-1 \ (\text{see [3.40]}).$$

D.2. If $F(0)=0$, $F(1)=0$, $\left.\frac{dF}{du}\right|_{u=0}=-\infty$, $\left.\frac{dF}{du}\right|_{u=1}=0$, say, $F=u\ln^2 u\,[b-\ln u\,(2+\ln u)]$, $b=\text{const}$, $\alpha=1$, then

$$\Theta=\begin{cases}\exp(-1/\xi) & \text{if } \xi\geqslant 0,\\ 0 & \text{if } \xi<0\end{cases}$$

(see p. 73 in [3.3]).

1.2. Semi-linear parabolic equations with variable coefficients:[13]

$$\varepsilon\frac{\partial u}{\partial t}-\varepsilon^2\frac{\partial}{\partial x}\left(\lambda(x,t)\frac{\partial u}{\partial x}\right)-\gamma(x,t)F(u)=0, \tag{3.8.0.5}$$

$$u(-\infty,t)=0,\quad u(\infty,t)=1.$$

The basic formulas for constructing asymptotic solutions:

$$u(x,t,\varepsilon)=[\Theta(\xi)+\varepsilon W_1(\xi,t)]|_{\xi=S/\varepsilon} \tag{3.8.0.6}$$

$$\begin{aligned}S(x,t,\varepsilon)&=(\beta(t)+\varepsilon\beta_3(t))(x+\varphi(t))\\&\quad+\beta_1(t)(x+\varphi(t))^2+\varepsilon S_1,\end{aligned} \tag{3.8.0.7}$$

$$\frac{b\,d\Theta}{d\xi}-\frac{d^2\Theta}{d\xi^2}-F(\Theta)=0,\ \Theta_{\xi\to-\infty}=0.\ \Theta|_{\xi\to+\infty}=1 \tag{3.8.0.8}$$

$$\Theta=\exp(l\xi)+o(\exp(l\xi)) \text{ as } \xi\to-\infty,$$

$$\Theta=1-\exp(-l_0\xi)+o(\exp(-l_0\xi)) \text{ as } \xi\to\infty,$$

$$\beta^2\lambda(-\varphi(t),t)=\gamma(-\varphi(t),t),\ \beta\frac{d\varphi}{dt}=\gamma(-\varphi(t),t)b, \tag{3.8.0.9}$$

$$\begin{aligned}\beta_1(t)&=-\left[l^2\beta\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi}-\frac{\partial\beta}{\partial t}\frac{l}{\beta}\right.\\&\quad\left.+\frac{1}{\beta}\left.\frac{\partial\gamma}{\partial x}\right|_{x=-\varphi}\right][2\lambda(-\varphi,t)B]^{-1},\end{aligned} \tag{3.8.0.10}$$

$$W_1(\xi,t)=C_1\frac{d\Theta}{d\xi}-\lambda^{-1}(-\varphi,t)\beta^{-2}(t)\frac{d\Theta}{d\xi}$$

---

[13] Here and below we assume that the variable coefficients in the equations are smooth functions of x and t and vanish nowhere.

$$\times \int_{-\infty}^{\xi} V\left(\frac{d\Theta}{d\xi'}\right)^{-2}\left(\int_{a}^{\xi'} fV^{-1}\frac{d\Theta}{d\mu}\,d\mu\right)d\xi', \qquad (3.8.0.11)$$

$$f = -\left\{\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}(\xi - S_1)\frac{d^2\Theta}{d\xi^2} + \beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}\frac{d\Theta}{d\xi}\right.$$
$$+ 2\beta_2\lambda(-\varphi, t)\frac{d\Theta}{d\xi} + 4\beta_1\lambda(-\varphi, t)(\xi - S_1)\frac{d^2\Theta}{d\xi^2}$$
$$\left.-\left[2\beta_1\frac{d\varphi}{dt} + \frac{d\beta}{dt}\right]\frac{\xi - S_1}{\beta}\frac{d\Theta}{d\xi} + \frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi}\frac{\xi - S_1}{\beta}F(\Theta)\right\}$$
$$+\left\{\left(\frac{\partial S_1}{\partial t} + \beta_3\frac{d\varphi}{dt}\right)\frac{d\Theta}{d\xi} - \lambda(-\varphi, t)\,2\beta\frac{d^2\Theta}{d\xi^2}\frac{\partial S_1}{\partial x}\right\},$$

$$V = \exp(b\xi), \qquad (3.8.0.12)$$

$$-\frac{\partial S_1}{\partial t} - \beta_3\frac{d\varphi}{dt} + \lambda(-\varphi, t)\,2\beta Yl + \beta l\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}$$
$$+ 2\beta_1\lambda(-\varphi, t) = 0, \qquad (3.8.0.13)$$

$$\int_{-\infty}^{\infty}\exp(b\xi)\left\{\left(\frac{\partial S_1}{\partial t} + \beta_3\frac{d\varphi}{dt}\right)\frac{d\Theta}{d\xi}\right.$$
$$-\lambda(-\varphi, t)\,2\beta\beta_3\frac{d^2\Theta}{d\xi^2} - \left[\beta(\xi - S_1)\frac{d^2\Theta}{d\xi^2}\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}\right.$$
$$+\beta\frac{d\Theta}{d\xi}\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi} - \left[2\beta_1\frac{d\varphi}{dt} + \frac{d\beta}{dt}\right]\frac{\xi - S_1}{\beta}\frac{d\Theta}{d\xi}$$
$$\left.\left.-F(\Theta)\frac{\xi - S_1}{\beta}\frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi}\right]\right\}\frac{d\Theta}{d\xi}\,d\xi = 0. \qquad (3.8.0.14)$$

1.2.A. KPP equations (see (3.8.0.2)). The asymptotic solution to problem (3.8.0.5) has the form (3.8.0.6). Function Θ can be found by solving problem (3.8.0.8), with $l = [b - (b^2 - b_{\min}^2)^{1/2}]/2$, $l_0 = -b/2 \mp (b^2/4 + |\, dF/d\Theta\, |_{\Theta=1}\,|)^{1/2}$, $\beta_3 \equiv 0$, and $S_1 = S_1(x, t)$ in (3.8.0.7). For functions β and φ we have system (3.8.0.9). Function β_1 is defined via formula (3.8.0.10), where $B = l(b^2 - b_{\min}^2)^{1/2}$, and $b > b_{\min} = 2(dF/d\Theta\,|_{\Theta=0})^{1/2}$. For function $S_1(x, t)$ we have Eq. (3.8.0.13), with $Y = \partial S_1/\partial x$. Function W_1 has the form (3.8.0.11), where $a = \mp\infty$ (see Remark, p. 320):

$$W_1 = O(\xi\exp\{2l\xi\}) \text{ as } \xi \to -\infty,\ b > (3/\sqrt{8})\,b_{\min};$$
$$W_1 = O(\xi^2\exp\{(b - l)\xi\}) \text{ as } \xi \to -\infty,$$
$$b = (3/\sqrt{8})\,b_{\min},\ b - l > l;$$
$$W_1 = O(\exp\{(b - l)\xi\}) \text{ as } \xi \to -\infty,$$
$$b_{\min} < b < (3/\sqrt{8})\,b_{\min};$$
$$W_1 = O(\xi^2\exp\{-l_0\xi\}) \text{ as } \xi \to \infty.$$

1.2.B. Zeldovich equations (see (3.8.0.3)). The asymptotic solution to problem (3.8.0.5) has the form (3.8.0.6). Function $\Theta(\xi)$ can be found by solving problem (3.8.0.8), with $b = b_0$ the Zeldovich constant, $l = b_0$, $l_0 = -b_0/2 + (b^2/4 + |dF/d\Theta|_{\Theta=1}|)^{1/2}$, and $S_1 \equiv S_1(t)$ in (3.8.0.7). For functions β and φ we have system (3.8.0.9). Function β_1 is defined via formula (3.8.0.10), where $B = b_0^2$. For functions $S_1(t)$ and $\beta_3(t)$ we have the system of equations (3.8.0.13), (3.8.0.14), with $Y = \beta_3$ and $l \equiv b \equiv b_0$. Function W_1 is defined in (3.8.0.11), where $a = -\infty$, while in (3.8.0.12),

$$W_1 = O(\xi \exp\{2b_0\xi\}) \text{ as } \xi \to -\infty;$$

$$W_1 = O(\xi^2 \exp\{-l_0\xi\}) \text{ as } \xi \to \infty; \quad \frac{\partial S_1}{\partial x} = \beta_3.$$

1.2.C. Semyonov equations (see (3.8.0.4)). The asymptotic solution to problem (3.8.0.5) has the form (3.8.0.6). Function Θ can be found by solving problem (3.8.0.8), with $l = b/2 + m$, $m = (b^2/4 + |dF/d\Theta|_{\Theta=0}|)^{1/2}$, $l_0 = -b/2 + (b^2/4 + |dF/d\Theta|_{\Theta=1}|)^{1/2}$, and $S_1 \equiv S_1(t)$ in (3.8.0.7). For functions β and φ we have system (3.8.0.9). Function β_1 is defined via formula (3.8.0.10), where $B = 2ml$. For functions $S_1(t)$ and $\beta_3(t)$ we have the system of equations (3.8.0.13), (3.8.0.14). Function W_1 is defined in (3.8.0.11), where $a = -\infty$, while in (3.8.0.12) $\partial S_1/\partial x \equiv \beta_3$ and $b < 2(dF/d\Theta|_{\Theta=a_1})^{1/2}$. Function W_1 has the following estimate:

$$W_1 = O(\xi \exp\{2l\xi\}) \text{ as } \xi \to -\infty;$$

$$W_1 = O(\xi^2 \exp\{-l_0\xi\}) \text{ as } \xi \to \infty.$$

1.3. Quasilinear parabolic equations:

$$\rho(u)\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left[K\left(u, \frac{\partial u}{\partial x}\right)\frac{\partial u}{\partial x}\right] - F(u) = 0.$$

The exact solutions to the equation

$$\rho(W_0)\frac{dW_0}{d\tau} - \frac{d}{d\tau}\left[K\left(W_0, \frac{dW_0}{d\tau}\right)\frac{dW_0}{d\tau}\right] - F(W_0) = 0$$

are related to the solution $\Theta(\xi)$ to the equation

$$\rho(\Theta)\frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} - F(\Theta)K\left(\Theta, \frac{d\Theta}{d\xi}\right) = 0 \qquad (3.8.0.14')$$

through the relationship

$$K\left(W_0, \frac{dW_0}{d\tau}\right)\frac{dW_0}{d\tau} = \frac{d\Theta}{d\xi}(\xi(W_0)),$$

where $\xi(W_0)$ is the inverse of function $\Theta(\xi)$ calculated at point $\Theta = W_0$.

1.4. *Quasilinear parabolic nonsingular equations*:

$$\varepsilon\rho(u)\,c(x,t)\frac{\partial u}{\partial t}-\varepsilon^2\frac{\partial}{\partial x}\left[K\left(u,\frac{\partial u}{\partial x}\right)\lambda(x,t)\frac{\partial u}{\partial x}\right]-\gamma(x,t)F(u)=0,$$

$$\rho(u)>0,\ K\left(u,\frac{\partial u}{\partial x}\right)>0, \qquad (3.8.0.15)$$

$$u(\infty,t)=1,\ u(-\infty,t)=0.$$

The basic formulas for constructing asymptotic solutions:

$$u(x,t,\varepsilon)=[W_0(\tau)+\varepsilon W_1(\tau,t)]|_{\tau=S/\varepsilon}, \qquad (3.8.0.15')$$

$$S(x,t,\varepsilon)=(\beta(t)+\varepsilon\beta_3(t))(x+\varphi(t))+\beta_1(x+\varphi)^2+\varepsilon S_1, \qquad (3.8.0.16)$$

$$b\rho(W_0)\frac{dW_0}{d\tau}-\frac{d}{d\tau}\left[K\left(W_0,\frac{dW_0}{d\tau}\right)\frac{dW_0}{d\tau}\right]-F(W_0)=0, \qquad (3.8.0.17)$$

$$W(\infty)=1,\quad W(-\infty)=0,$$

$$\beta^2\lambda(-\varphi,t)=\gamma(-\varphi(t),t), \qquad (3.8.0.18)$$

$$c(-\varphi(t),t)\,\beta\frac{d\varphi(t)}{dt}=b\gamma(-\varphi(t),t),$$

$$\beta_1(t)=-\left[\beta l^2K(0,0)\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}-\frac{l}{\beta}\frac{d\beta}{dt}c(-\varphi,t)\ \rho(0)\right.$$
$$\left.-\beta\frac{d\varphi}{dt}\frac{\partial c}{\partial x}\Big|_{x=-\varphi}\rho(0)\,l+\frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi}\frac{1}{\beta}\right][2\lambda(-\varphi,t)B]^{-1},$$
$$(3.8.0.19)$$

$$W_1(\tau,t)=C_1\frac{dW_0}{d\tau}-\lambda^{-1}(-\varphi,t)\beta^{-2}(t)\frac{dW_0}{d\tau}\int_{-\infty}^{\tau}V\left(\frac{dW_0}{d\tau'}\right)^{-2}\left(\int_a^{\tau'}f\frac{dW_0}{d\tau}\right.$$
$$\left.\times\left[V\left(K\left(W_0,\frac{dW_0}{d\tau}\right)\right)+\frac{dW_0}{d\tau}\frac{\partial K(W_0,\mu)}{\partial\mu}\right]^{-1}d\tau\right)d\tau',$$
$$(3.8.0.20)$$

$$f=-\left\{\beta(\tau-S_1)\frac{d\Pi}{d\tau}+\beta\Pi\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}+2\beta_1\Pi\lambda(-\varphi,t)\right.$$
$$+4\beta_1\lambda(-\varphi,t)(\tau-S_1)\frac{d\Pi}{d\tau}-\beta\frac{d\varphi}{dt}\frac{\partial c}{\partial x}\Big|_{x=-\varphi}\rho(W_0)\frac{dW_0}{d\tau}\frac{\tau-S_1}{\beta}$$
$$-\left[2\beta_1\frac{d\varphi}{dt}+\frac{d\beta}{dt}\right]\frac{\tau-S_1}{\beta}\frac{dW_0}{d\tau}c(-\varphi,t)\rho(W_0)$$
$$\left.+\frac{\partial\gamma}{\partial x}\Big|_{x=-\varphi}\frac{\tau-S_1}{\beta}F(W_0)\right\}$$
$$+\left\{\left(\frac{\partial S_1}{\partial t}+\beta_3\frac{d\varphi}{dt}\right)\frac{dW_0}{d\tau}c(-\varphi,t)\rho(W_0)\right.$$
$$\left.-\lambda(-\varphi,t)\,2\beta\frac{\partial S_1}{\partial x}\frac{d\Pi}{d\tau}\right\}, \qquad (3.8.0.21)$$

$$V = K^{-2}\left(W_0,\ \frac{dW_0}{d\tau}\right)\left[\int \left\{b\rho(W_0) - \frac{\partial}{\partial\tau}\left(\frac{dW_0}{d\tau}\frac{\partial K}{\partial\mu}\ (W_0,\ \mu)\right)\right\}\right.$$
$$\left. K^{-1}\left(W_0,\ \frac{dW_0}{d\tau}\right)d\tau\right],$$

$$\Pi = K\left(W_0,\ \frac{dW_0}{d\tau}\right)\frac{dW_0}{d\tau},$$

$$\left(-\frac{\partial S_1}{\partial t} - \beta_3\frac{d\varphi}{dt}\right)\rho(0)c(-\varphi,\ t) + \lambda(-\varphi,\ t)\,l_2\beta K(0,\ 0)Y$$
$$+\beta K(0,\ 0)\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi} + 2\beta_1\lambda(-\varphi,\ t)K(0,\ 0) = 0, \qquad (3.8.0.22)$$

$$\int_{-\infty}^{\infty} M\left\{\left(\frac{\partial S_1}{\partial t} + \beta_3\frac{d\varphi}{dt}\right)\rho(W_0)c(-\varphi,\ t)\Pi\frac{dW_0}{d\tau} - 2\lambda(-\varphi,\ t)\,\beta\beta_3\Pi\frac{d\Pi}{d\tau}\right.$$
$$-\beta\Pi\frac{d\Pi}{d\tau}(\tau - S_1) - \left[\beta\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi} + 2\beta_1\lambda(-\varphi,\ t)\right]\Pi^2$$
$$-4\beta_1\lambda(-\varphi,\ t)\Pi\frac{d\Pi}{d\tau}(\tau - S_1) + \beta\frac{d\varphi}{dt}\rho(W_0)\Pi\frac{\tau - S_1}{\beta}\frac{dW_0}{d\tau}$$
$$+\left(2\beta_1\frac{d\varphi}{dt} + \frac{d\beta}{dt}\right)\frac{1}{\beta}\Pi(\tau - S_1)\frac{dW_0}{d\tau}$$
$$\left.-\frac{1}{\beta}\Pi F(W_0)(\tau - S_1)\left.\frac{\partial\gamma}{\partial x}\right|_{x=-\varphi}\right\}d\tau = 0, \qquad (3.8.0.23)$$

$$M = \exp\left[-\int\left\{b\rho(W_0)K^{-1}\left(W_0,\ \frac{dW_0}{d\tau}\right)\right.\right.$$
$$\left.\left.-K^{-1}\left(W_0,\ \frac{dW_0}{d\tau}\right)\frac{\partial}{\partial\tau}\left(\frac{\partial W_0}{\partial\tau}\frac{\partial K(W_0,\ \mu)}{\partial\mu}\right)\right\}d\tau\right].$$

1.4.A. Equation (3.8.0.15) with F satisfying (3.8.0.2). The asymptotic solution to problem (3.8.0.15) has the form (3.8.0.15′). For function $W_0(\tau)$ we have problem (3.8.0.17), with $\beta_3 \equiv 0$ and $S_1 \equiv S_1(x,\ t)$ in (3.8.0.16). For functions β and φ we have system (3.8.0.18). Function β_1 is defined via (3.8.0.19), where

$$B = 2l\,[b^2\rho^2(0)/4 - K(0,\ 0)(dF/dW_0|_{W_0=0})]^{1/2},$$
$$b > b_{\min} = (2/\rho(0))\,[K(0,\ 0)(dF/dW_0|_{W_0=0})]^{1/2},$$
$$l = (b - \sqrt{b^2 - b_{\min}^2})\,\rho(0)\,(2K(0,\ 0))^{-1}.$$

For function $S_1(x,\ t)$ we have Eq. (3.8.0.22), with $Y = \beta_3$. Finally, function W_1 has the form (3.8.0.20), where $a = +\infty$ (see p. 320).

1.4.B. Equation (3.8.0.15) with F satisfying (3.8.0.3). The asymptotic solution to problem (3.8.0.15) has the form (3.8.0.15′). For function $W_0(\tau)$ we have problem (3.8.0.17), with $b = b_0^*$ the Zeldo-

vich constant for Eq. (3.8.0.14′), and $S_1 \equiv S_1(t)$ in (3.8.0.16). For functions β and φ we have system (3.8.0.18). Function β_1 is defined via (3.8.0.19), where

$$B=(b_0\rho(0))^2K^{-1}(0,\ 0),\quad l=b_0\rho(0)/K(0,\ 0).$$

For functions $S_1(t)$ and $\beta_3(t)$ we have the system of equations (3.8.0.22), (3.8.0.23), with $Y=\beta_3$. Function W_1 and Wronskian V are defined via (3.8.0.20) and (3.8.0.21), with $a=-\infty$.

1.4.C. Equation (3.8.0.15) with F satisfying (3.8.0.4). The asymptotic solution to problem (3.8.0.15) has the form (3.8.0.15′). For function W_0 we have problem (3.8.0.17), with $S_1 \equiv S_1(t)$ in (3.8.0.6). For functions β and φ we have system (3.8.0.18). Function β_1 is defined via (3.8.0.19), where

$$B=2ml,\ l=[l\rho(0)/2+m]\,K^{-1}(0,\ 0),$$
$$m=[b^2\rho^2(0)/4+K(0,\ 0)\,dF/dW_0|_{W_0=0}]^{1/2}.$$

For functions $S_1(t)$ and $\beta_3(t)$ we have the system of equations (3.8.0.22), (3.8.0.23), with $Y=\beta_3$. For function W_1 we have (3.8.0.20), with $a=-\infty$.

1.4.D. The asymptotic solution to the boundary value problem for the KPP equation with a variable root of the equation $\mathcal{F}(x,\ t,\ u)=0$, that is,

$$\varepsilon\frac{\partial u}{\partial t}-\varepsilon^2\frac{\partial}{\partial x}\left(\lambda(x,\ t)\frac{\partial u}{\partial x}\right)-\gamma(x,\ t)\,u\,(\mu(x)-u)=0,$$
$$u|_{x\to+\infty}=\mu(x)|_{x\to\infty},\ u|_{x\to-\infty}=0,\ 0<\mu<1,$$

has the form

$$u(x,\ t,\ \varepsilon)=\mu(x)\,[\Theta(S/\varepsilon)+\varepsilon W_1(S/\varepsilon,\ t)],$$

where function S has the form (3.8.0.7), $\beta_3\equiv 0$, $S_1\equiv S_1(x,\ t)$, function $\Theta(\xi)$ solves problem (3.8.0.8), and functions $\beta(t)$ and $\varphi(t)$ are found by solving the following system of equations:

$$\beta\frac{d\varphi}{dt}=b\gamma(-\varphi,\ t)\,\mu(-\varphi),\ \lambda(-\varphi,\ t)\,\beta^2=\gamma(-\varphi,\ t)\,\mu(-\varphi).$$

Function β_1 has the form

$$\beta_1=-\Big[\beta l^2\frac{\partial(\lambda\mu)}{\partial x}\Big|_{x=-\varphi}-\mu(-\varphi)\frac{l}{\beta}\frac{d\beta}{dt}$$
$$+\frac{1}{\beta}\frac{\partial(\gamma\mu^2)}{\partial x}\Big|_{x=-\varphi}-l\frac{d\varphi}{dt}\frac{\partial\mu}{\partial x}\Big|_{x=-\varphi}\Big]\,[2\lambda(-\varphi,\ t)\,\mu(-\varphi)\,b]^{-1},$$

where $B = l\,(b^2 - b_{\min}^2)^{1/2}$, $l = [b - (b^2 - 4)^{1/2}]/2$, and $b_{\min} = 2$ (see Subsection 1.4.A). Function S_1 is found from the equation

$$-\frac{\partial S_1}{\partial t}\mu(-\varphi)\,l + 2\lambda(-\varphi, t)\,\mu(-\varphi)\,l^2\beta\frac{\partial S_1}{\partial x} + \mu(-\varphi)\,l\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}$$
$$+ 2\lambda(-\varphi, t)\frac{\partial\mu}{\partial x}\Big|_{x=-\varphi} l\beta + 2\beta_1\lambda(-\varphi, t)\,\mu(-\varphi)\,l = 0.$$

Function W_1 has the form (3.8.0.11), with $a = +\infty$, and f is defined by the formula

$$f = \Big\{\mu(-\varphi)\left(2\beta_1\frac{d\varphi}{dt} + \frac{d\beta}{dt}\right)\frac{\xi - S_1}{\beta}\frac{d\Theta}{d\xi}$$
$$+ \frac{d\mu}{dx}\Big|_{x=-\varphi}\frac{\xi - S_1}{\beta}\beta\frac{d\varphi}{dt}\frac{d\Theta}{d\xi} - \mu(-\varphi)\beta\frac{d\Theta}{d\xi}\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}$$
$$- 2\lambda(-\varphi, t)\frac{d\mu}{dx}\Big|_{x=-\varphi}\beta\frac{d\Theta}{d\xi}$$
$$- \beta(\xi - S_1)\frac{d^2\Theta}{d\xi^2}\frac{\partial(\lambda\mu)}{\partial x}\Big|_{x=-\varphi} - 2\beta_1\lambda(-\varphi, t)\,\mu(-\varphi)\frac{d\Theta}{d\xi}$$
$$- 4\beta_1(t)\,\lambda(-\varphi, t)\,\mu(-\varphi)(\xi - S_1)\frac{d^2\Theta}{d\xi^2}$$
$$- \frac{\xi - S_1}{\beta}\Theta(1 - \Theta)\frac{\partial(\gamma\mu^2)}{\partial x}\Big|_{x=-\varphi}\Big\}$$
$$- \left\{\frac{d\Theta}{d\xi}\frac{\partial S_1}{\partial t}\mu(-\varphi) - 2\lambda(-\varphi, t)\,\mu(-\varphi)\beta\frac{\partial S_1}{\partial x}\frac{d^2\Theta}{d\xi^2}\right\},\quad V = \exp(b\xi).$$

The asymptotic solution to a problem involving the Zeldovich or Semyonov equation with a variable root $\mu = \mu(x, t)$ can be constructed by reasoning along similar lines (see p. 87 in [3.3]). Another asymptotic solution to the semi-linear parabolic equation is given in [3.30].

1.5. *Hyperbolic quasilinear nonsingular equation* (for the one-dimensional case see [3.3], while the multi-dimensional case has been considered in Sections 3.6.5 and 3.6.6).

The boundary value problem:

$$\varepsilon\frac{\partial u}{\partial t} + \varepsilon^2\frac{\partial^2 u}{\partial t^2} - \varepsilon^2\frac{\partial}{\partial x}\left(\lambda(x, t)\,K(u)\frac{\partial u}{\partial x}\right) - \mathcal{F}(u) = 0, \tag{3.8.0.24}$$
$$\mathcal{F}(a_i) = 0,\ i = 0, 1,\ K(u) > 0,\ u|_{x\to+\infty} = a_1,\ u|_{x\to-\infty} = a_0.$$

The basic formulas for constructing asymptotic solutions:

$$u(x, t, \varepsilon) = \chi(\tau) + \varepsilon W_1(\tau, t), \tag{3.8.0.25}$$
$$S(x, t, \varepsilon) = \beta(t)(x + \varphi(t)) + \beta_1(t)(x + \varphi(t))^2 + \varepsilon S_1(x, t) \tag{3.8.0.26}$$
$$b\frac{d\chi}{d\tau} - \mu\frac{d}{d\tau}\left[|K(\chi) - b^2|\frac{d\chi}{d\tau}\right] - \mathcal{F}(\chi) = 0, \tag{3.8.0.27}$$

$$\chi(-\infty) = a_0, \quad \chi(+\infty) = a_1,$$

$$\beta^2(t)\,\lambda(-\varphi,\ t) = 1, \quad \beta\frac{d\varphi}{dt} = b, \tag{3.8.0.28}$$

$$\beta_1(t) = \Big\{-\frac{\mu}{\beta}\frac{d\beta}{dt} - 2l\frac{d\varphi}{dt}\frac{d\beta}{dt}\,|K(a_0) - b^2|^{-1} \tag{3.8.0.29}$$

$$+\beta l K(a_0)\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi}|K(a_0) - b^2|^{-1}\Big\}\{-2\lambda(-\varphi,\ t)(b + 2l\mu)\}^{-1},$$

$$-\mu\frac{\partial S_1}{\partial t} - \frac{2l\beta}{|K(a_0) - b^2|}\frac{d\varphi}{dt}\frac{\partial S}{\partial x} - \mu\Big[\frac{d}{dt}\left(\beta\frac{d\varphi}{dt}\right)$$

$$+2\beta_1\left(\frac{d\varphi}{dt}\right)^2\Big] + \mu K(a_0)\beta\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi}$$

$$+\frac{2\lambda(-\varphi,\ t)\,lK(a_0)}{|K(a_0) - b^2|}\beta\frac{\partial S_1}{\partial x} = 0,$$

$$W_1(\tau,\ t) = C_1\frac{d\chi}{d\tau} + \lambda^{-1}(-\varphi,\ t)\,\beta^{-2}(t)\frac{d\chi}{d\tau}$$

$$\times\int_{-\infty}^{\tau} V\left(\frac{d\chi}{d\tau'}\right)^{-2}\left(\int_{+\infty}^{\tau'} f\frac{d\chi}{d\tau}V^{-1}(K(\chi) - b^2)^{-1}\,d\tau\right)d\tau', \tag{3.8.0.30}$$

$$f = -\frac{d\chi}{d\tau}\Big[\frac{\tau - S_1}{\beta}\left(\frac{d\beta}{dt} + 2\beta_1\frac{d\varphi}{dt}\right) + \frac{\partial S_1}{\partial t}\Big]$$

$$-\frac{d^2\chi}{d\tau^2}\Big[2\beta\frac{d\varphi}{dt}\frac{\partial S_1}{\partial t} + 2(\tau - S_1)\frac{d\varphi}{dt}\left(\frac{d\beta}{dt} + 2\beta_1\frac{d\varphi}{dt}\right)\Big] \tag{3.8.0.31}$$

$$-\frac{d\chi}{dt}\Big[\frac{d}{dt}\left(\beta\frac{d\varphi}{dt}\right) + 2\beta_1\left(\frac{d\varphi}{dt}\right)^2\Big] + \beta(\tau - S_1)\frac{d\Pi}{d\tau}\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi}$$

$$+\beta\Pi\left.\frac{\partial\lambda}{\partial x}\right|_{x=-\varphi} + 2\beta_1\lambda(-\varphi,\ t)\,\Pi$$

$$+4\beta_1\lambda(-\varphi,\ t)(\tau - S_1)\frac{d\Pi}{d\tau} + \lambda(-\varphi,\ t)\,2\beta\frac{\partial S_1}{\partial x}\frac{d\Pi}{d\tau},$$

$$V = (K(\chi) - b^2)^{-2}\exp\left(b\int_0^{\tau}[K(\chi) - b^2]^{-1}\,d\tau\right), \quad \Pi = K(\chi)\frac{d\chi}{d\tau}. \tag{3.8.0.32}$$

1.5.A. The asymptotic solution to problem (3.8.0.24) with

$$\left.\frac{d\mathcal{F}}{du}\right|_{u=a_0} > 0, \quad \left.\frac{d\mathcal{F}}{du}\right|_{u=a_1} < 0, \quad K(\chi) - b^2 > 0$$

for $\chi \in [a_0,\ a_1]$ has the form (3.8.0.25). For function χ we have problem (3.8.0.27), with

$$b > b_{\min} = 2\left\{\left.\frac{d\mathcal{F}}{d\chi}\right|_{\chi=a_0}K(a_0)\Big[1 + 4\left.\frac{d\mathcal{F}}{d\chi}\right|_{\chi=a_0}\Big]^{-1}\right\}^{1/2}.$$

For functions β and φ we have system (3.8.0.28). Function β_1 is defined via (3.8.0.29), with $\mu = 1$ and $l = b/2 - \{(b^2 - b^2_{\min}) \times [1 + 4\,(d\mathcal{F}/d\chi\,|_{\chi=a_0})]\}^{1/2}\,2$. For function S_1 we have Eq. (3.8.0.30). Finally, function W_1 is defined via (3.8.0.31), (3.8.0.32).

1.5.B. The asymptotic solution to problem (3.8.0.24) with

$$\frac{d\mathcal{F}}{du}\bigg|_{u=a_0} > 0,\ \frac{d\mathcal{F}}{du}\bigg|_{u=a_1} < 0,\ K(\chi) - b^2 < 0$$

for $\chi \in [a_0, a_1]$ has the form (3.8.0.25). For function χ we have problem (3.8.0.27), with

$$b < b_{\max} = -[|\,d\mathcal{F}/d\chi|_{\chi=a_1}|\ \ K(a_1)\,(4\,|\,d\mathcal{F}/d\chi|_{\chi=a_1}| - 1)^{-1}]^{1/2}.$$

For functions β and φ we have system (3.8.0.28). Function β_1 is defined via (3.8.0.29), with $\mu = -1$ and

$$l = b/2 + [b^2/4 + (d\mathcal{F}/d\chi|_{\chi=a_0})\,|K(a_0) - b^2|]^{1/2}.$$

For function S_1 we have Eq. (3.8.0.30), and function W_1 is defined via (3.8.0.31) and (3.8.0.32).

The asymptotic solution to problem (3.8.0.24) involving equations with the function $\mathcal{F}$ satisfying (3.8.0.3) or (3.8.0.4) can be built by reasoning along similar lines (see [3.3]).

1.6. *Quasilinear singular parabolic equations*:

$$\begin{aligned}
&\varepsilon\frac{\partial u}{\partial t} - \varepsilon^2\frac{\partial}{\partial x}\left(K(u)\frac{\partial u}{\partial x}\right) - F(u) = 0,\\
&K(0) = 0,\ K(u) > 0 \text{ for } u > 0,\\
&u(-\infty,\ t) = 0,\ u(\infty,\ t) = 1.
\end{aligned}\tag{3.8.0.33}$$

1.6.A. Characteristic exact solutions to Eq. (3.8.0.33):

A.1. $K(u) = (2-q)\,u^{1-q}$, $\mathcal{F} = \dfrac{mu^q}{(1-q)^2}(1 - u^{2(1-q)})$

$$\times\left[\frac{(1-q)\,b - 2 + q}{m} + u^{2(1-q)}\right],$$

$$0 < q < 1,\ b > \frac{2-q}{1-q},\ m = (3 - 2q)(2 - q),$$

$$u = \begin{cases}\left[\tan\left(\dfrac{x + bt - x_0}{\varepsilon}\right)\right]^{1/(1-q)} & \text{for } x - x_0 + bt > 0,\\ 0 \text{ for } x - x_0 + bt \leqslant 0\end{cases}$$

(see p. 33 in [3.3]).

A.2 $K = Du^m$, $F = u^q(1 - u^\nu)$, $m,\ q > 0$, $m + q = 1$, $\nu > 0$,

$$u = \chi(\xi)|_{\xi=(\alpha x + bt)/\varepsilon},\ \alpha = 1,\ \ b = (4+\nu)\,\nu/(4 + 2\nu),$$

$D = \nu^2/(4 + 2\nu)$ (see p. 78 in [3.3]). Function $\chi(\xi)$ is given by the integral

$$\int_0^{\chi} \frac{\chi^{m-1} d\chi}{1-\chi^{\nu/2}} = \frac{\nu\xi}{2}.$$

A.3. $K(u) = 2u$, $F = u[1 - 2u(1 + 2\ln u)\ln u]\ln^2 u$,

$$u = \chi(\xi)|_{\xi=(x+t)/\varepsilon}, \quad \chi = \begin{cases} \exp(-\tau^{-1}) & \text{if } \tau \geqslant 0, \\ 0 & \text{if } \tau < 0. \end{cases}$$

A.4. The asymptotic localized solution to Eq. (3.8.0.33) (the "product" of exact solutions):

$$u = \left[W\left(\frac{x+bt}{\varepsilon}\right)(1-E_1) + E_1\right]\left[W\left(\frac{-x+bt+\gamma_1}{\varepsilon}\right)(1-E_2) + E_2\right],$$

where $W(\xi)$ is an exact solution to problem (3.8.0.33), and E_1 and E_2 are infinitely differentiable functions,

$$E_1 = \begin{cases} 0 & \text{if } x < -bt - a_0, \\ 1 & \text{if } x > -bt - a_0 + \delta_1, \quad 0 < \delta_1 < 1, \end{cases}$$

$$F_2 = \begin{cases} 0 & \text{if } x > bt + \gamma_1 - a_0, \\ 1 & \text{if } x < bt + \gamma_1 - a_0 - \delta_1 \end{cases}$$

(see [3.3] and Section 3.7).

1.7. Quasilinear parabolic singular equations with variable coefficients:

A.1. The asymptotic localized solution to the boundary value problem

$$\varepsilon \frac{\partial u}{\partial t} - \varepsilon^2 \frac{\partial}{\partial x}\left(\lambda(x, t) K(u) \frac{\partial u}{\partial x}\right) - F(u, x, t) = 0, \tag{3.8.0.34}$$

$$K(0) = 0, \ F(0) = 0, \ F(1) = 0,$$
$$K(u) \sim u^{k-1} \text{ and } F(u) \sim \gamma^2(x, t) u^q \text{ as } u \to 0,$$
$$u(-\infty, t) = 0, \ u(\infty, t) = 1,$$

has the form

$$u = [W_0(S/\varepsilon) + \varepsilon W_1(S/\varepsilon, t)].$$

Let us put $K(u) = u^{k-1}\rho(u)$ and $F = \gamma^2(x, t) u^q G(u)$, with $\rho(0) \neq 0$, $G(0) \neq 0$, and $G(1) = 0$. Function $W_0(\tau)$ satisfies the equation

$$b\frac{dW_0}{d\tau} - \frac{d}{d\tau}\left(K(W_0)\frac{dW_0}{d\tau}\right) - F(W_0) = 0,$$
$$W_0(\tau)|_{\tau=0} = 0, \ W_0(\tau)|_{\tau\to\infty} \to 1 - 0.$$

Function $S(x, t, \varepsilon)$ has the form[14]

$S(x, t, \varepsilon) = \beta(t, \varepsilon)(x + \varphi(t, \varepsilon)) + \beta_1(t, \varepsilon)(x + \varphi(t, \varepsilon))^2$,

where functions β and φ are defined via the system of equations.

$$\beta\sqrt{\lambda(-\varphi, t)} = \gamma(-\varphi, t), \ \beta\frac{d\varphi}{dt} = \gamma^2(-\varphi, t)\,b.$$

For function W_0 the following estimates hold true:

$$W_0(\tau) = O(\tau^\alpha) \text{ as } \tau \to 0,$$

$$W_0 = 1 - \exp\left(-\frac{|l_1|\tau}{\rho(1)}\right) + o\left(\exp\left(-\frac{|l_1|\tau}{\rho(1)}\right)\right) \text{ as } \tau \to \infty.$$

Here

(a) if $k + q > 2$, $q \geqslant 1$, and $F(u) > 0$ for $u \in (0, 1)$, then $\alpha = 1/(k-1)$ and $l_1 = -b_0/2 + [b_0^2/4 + dR/d\Theta\,|_{\Theta=1}]^{1/2}$, with b_0 the Zeldovich constant in the Zeldovich equation

$$b_0\frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} - R(\Theta) = 0;$$

(b) if $k + q = 2$, $q < 1$, and $F(u) > 0$ for $u \in (0, 1)$, then $\alpha = (k-1)^{-1} = (1-q)^{-1}$, $l = -b/2 + [b^2/4 - dR/d\Theta\,|_{\Theta=1}]^{1/2}$ and $b > 2\,(dR/d\Theta\,|_{\Theta=0})^{1/2}$;

(c) if $k + q > 2$, $q < 1$, and $F(u) < 0$ for $u \in (0, 1)$, then $\alpha = (1-q)^{-1}$, $l_1 = b/2 - [b^2/4 - dR/d\Theta\,|_{\Theta=1}]^{1/2}$, and $b \leqslant -2(dR/d\Theta\,|_{\Theta=1})^{1/2}$, where $R(\Theta) = \rho(\Theta)\,\Theta^{k+q-1}G(\Theta) = K(\Theta)\,F(\Theta)$.

Function $W_1(\tau, t)$ has the form

$$W_1 = -\lambda^{-1}(-\varphi, t)\,\beta^{-2}(t, \varepsilon)\,\frac{\partial W_0}{\partial\tau} \times \int_0^\tau \left(\frac{V}{(\partial W_0/\partial\tau')^2}\int_a^{\tau'} \frac{f(t, \xi)\,(\partial W_0/\partial\xi)}{VK(W_0)}\,d\xi\right)d\tau'. \tag{3.8.0.35}$$

Here

(a) if $q \geqslant 1$, $k + q > 2$, then the lower limit a in the inner integral in (3.8.0.35) vanishes;

(b) if $q < 1$ and $k + q = 2$, then $a = 0$;

(c) if $q < 1$ and $k + q > 2$, then $a = +\infty$.

Functions V and f are defined thus:

$$V = K^{-2}(W_0)\exp\left(b\int K^{-1}\,d\tau\right),$$

[14] Here and below, if functions β, φ, and β_1 are independent of ε at $t = 0$, we must put $\beta(t, 0)$, $\varphi(t, 0)$, and $\beta_1(t, 0)$ in all formulas.

$$f=\Big\{2\beta_1\lambda(-\varphi,\ t)\,K(W_0)\frac{dW_0}{d\tau}+4\beta_1\tau\frac{d}{d\tau}\left(K(W_0)\frac{dW_0}{d\tau}\right)\lambda(-\varphi,\ t)$$

$$-\left[2\beta_1\frac{d\varphi}{dt}+\frac{d\beta}{dt}\right]\frac{\tau}{\beta}\frac{dW_0}{d\tau}+\frac{\partial\gamma^2(x,\ t)}{\partial x}\Big|_{x=-\varphi}\frac{\tau}{\beta}F(W_0)$$

$$+\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}\tau\frac{d}{d\tau}\left(K(W_0)\frac{dW_0}{d\tau}\right)$$

$$+\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}K(W_0)\frac{dW_0}{d\tau}\Big\}.$$

The following estimates hold true:

$$W_1=O(\tau^{\alpha+1})\ \text{as}\ \tau\to 0;$$

$$W_1=O\left(\tau^2\exp\left(-\frac{|l_1|\,\tau}{K(1)}\right)\right)\ \text{as}\ \tau\to\infty.$$

Function β_1 is specified by the condition

$$(i)\qquad \int_0^\infty f(t,\ \tau)\frac{dW_0}{d\tau}V^{-1}K^{-1}(W_0)\,d\tau=0$$

if $k+q>2$ and $q\geqslant 1$
or if $k+q=2$ and $q<1$
or by the condition

$$(ii)\ \lim_{\tau\to 0}\left\{\int_0^\tau V\left(\frac{dW_0}{d\tau}\right)^{-2}\left(\int_{+\infty}^{\tau'} f\frac{dW_0}{d\xi}V^{-1}K^{-1}(W_0(\xi))\,d\xi\right)d\tau'\right\}=0^{15}$$

if $k+q>2$ and $q<1$, and is given by the following formulas:
(1) if $k+q>2$, $q\geqslant 1$,
or if $k+q=2$, $q<1$, then

$$\beta_1(t,\ \varepsilon)=\left\{2I_2\lambda(-\varphi,\ t)+4I_0\lambda(-\varphi,\ t)-\frac{2}{\beta}\frac{d\beta}{dt}I_1\right\}^{-1}$$

$$\times\left\{-\beta I_0\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}-\beta I_2\frac{d\lambda}{dx}\Big|_{x=-\varphi}+\frac{1}{\beta}\frac{d\beta}{dt}I_1-\frac{1}{\beta}I_3\frac{\partial\gamma^2}{\partial x}\Big|_{x=-\varphi}\right\},$$

$$I_0=\int_0^\infty N\tau\frac{d}{d\tau}\left(K(W_0)\frac{dW_0}{d\tau}\right)d\tau,\quad I_2=\int_0^\infty NK(W_0)\frac{dW_0}{d\tau}\,d\tau,$$

$$I_1=\int_0^\infty N\tau\frac{dW_0}{d\tau}\,d\tau,\quad I_3=\int_0^\infty N\tau F(W_0)\,d\tau,$$

[15] The condition (*ii*) can be rewritten in the form of (*i*).

$$N = K(W_0)\frac{dW_0}{d\tau}\exp\left(-b\int\frac{d\tau}{K(W_0)}\right);$$

(2) if $k+q>2$, $q<1$, then

$$\beta_1(t, \varepsilon) = \lim_{\tau\to 0}\Big[\Big\{2M_2\lambda(-\varphi, t)+4M_0\lambda(-\varphi, t)-\frac{2}{\beta}\frac{d\varphi}{dt}M_1\Big\}^{-1}$$
$$\times\Big\{-\beta M_0\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}+\frac{M_1}{\beta}\frac{d\beta}{dt}-\beta M_2\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}$$
$$-\frac{M_3}{\beta}\frac{\partial\gamma^2}{\partial x}\Big|_{x=-\varphi}\Big\}\Big],$$

where

$$M_0=\int_0^\tau V\left(\frac{dW_0}{d\tau}\right)^{-2}\left(\int_\infty^\tau N(\tau')\tau'\frac{d}{d\tau'}\left(K(W_0)\frac{dW_0}{d\tau'}\right)d\tau'\right)d\tau,$$

$$M_1=\int_0^\tau V\left(\frac{dW_0}{d\tau}\right)^{-2}\left(\int_\infty^\tau N(\tau')\tau'\frac{dW_0}{d\tau'}d\tau'\right)d\tau,$$

$$M_2=\int_0^\tau V\left(\frac{dW_0}{d\tau}\right)^{-2}\left(\int_\infty^\tau N(\tau')K(W_0)\frac{dW_0}{d\tau'}d\tau'\right)d\tau,$$

$$M_3=\int_0^\tau V\left(\frac{dW_0}{d\tau}\right)^{-2}\left(\int_\infty^\tau N(\tau')\tau' F(W_0)\,d\tau'\right)d\tau.$$

For the multidimensional case see [3.3].

1.8. The equation

$$\begin{aligned}\varepsilon\frac{\partial u}{\partial t}-\varepsilon^2\frac{\partial}{\partial x}\left(\lambda(x, t)\frac{\partial u^k}{\partial x}\right)-\delta(x, t)\frac{\partial u}{\partial x}&\\ -\gamma^2(x, t)u^q[A(x, t)-u^\mu]\nu=0,&\qquad(3.8.0.36)\\ u(-\infty, t)=0,\ u(\infty, t)=1&\end{aligned}$$

with a variable root of the equation $\mathscr{F}(x, t, u)=0$.

The asymptotic localized solution to it has the form

$$\begin{aligned}u(x, t, \varepsilon)&=W([S/\varepsilon+\varepsilon g(S/\varepsilon, t, \varepsilon)+O(\varepsilon^2)], t, \varepsilon)\\ &=A^{1/\mu}(x, t)\chi(S/\varepsilon+\varepsilon g(S/\varepsilon, t, \varepsilon)+O(\varepsilon^2))\end{aligned}$$

if $k>1$, $q>0$, $k+q\geqslant 2$, $\mu>0$, and function $A(x, t)$ satisfies

$$\frac{\partial A}{\partial t}-\delta(x, t)\frac{\partial A}{\partial x}=0, \qquad (3.8.0.37)$$

which is the limiting equation as $\varepsilon \to 0$, with $u^\mu \to A$. The function $\chi(\xi)$ solves the boundary value problem

$$b\frac{d\chi}{d\xi} - \frac{d^2\chi^k}{d\xi^2} - \nu\chi^q(1-\chi^\mu) = 0,\ \chi|_{\xi=0} = 0,\ \chi|_{\xi\to\infty} = 1,\ \frac{d\chi^k}{d\xi}\Big|_{\xi=0} = 0. \tag{3.8.0.38}$$

Function $S(x, t, \varepsilon)$ has the form

$$S(x, t, \varepsilon) = \beta(t, \varepsilon)(x + \varphi(t, \varepsilon)) + \beta_1(t, \varepsilon)(x + \varphi(t, \varepsilon))^2,$$

where functions β and φ can be found by solving the system of equations

$$\begin{gathered}\beta A^{(1-q)/\mu-1}(-\varphi, t)\left(\frac{d\varphi}{dt} - \delta(-\varphi, t)\right) = b\gamma^2(-\varphi, t),\\ \beta^2\lambda(-\varphi, t)A^{(k-q)/\mu-1}(-\varphi, t) = \gamma^2(-\varphi, t).\end{gathered} \tag{3.8.0.39}$$

For function $\chi(\tau)$ the following estimates hold true:

$$\chi = C_1\xi^\alpha - C_2\xi^{\alpha+\beta} + o(\xi^{\alpha+\beta}) \text{ as } \xi \to 0,$$
$$\chi = 1 - \exp(-l_0\xi/k) + o(\exp(-l_0\xi/k)) \text{ as } \xi \to \infty.$$

(a) If $k+q > 2$, $q \geqslant 1$, and $\nu = 1$, then $\alpha = (k-1)^{-1}$ and $\beta = 1 + (q-1)(k-1)^{-1}$ and $l_0 = -b_0/2 + [b_0^2/4 - (dR/d\Theta|_{\Theta=1})]^{1/2}$, where b_0 is the Zeldovich constant in the equation

$$b_0\frac{d\Theta}{d\xi} - \frac{d^2\Theta}{d\xi^2} - R(\Theta) = 0;$$

(b) if $k+q = 2$, $q < 1$, and $\nu = 1$, then $\alpha = (k-1)^{-1} = (1-q)^{-1}$, $l_0 = -b/2 + [b^2/4 - (dR/d\Theta/_{\Theta=1})]^{1/2}$, $b > 2(dR/d\Theta|_{\Theta=0})^{1/2} = 2\sqrt{k}$,

$$\beta = \frac{\alpha kz(k+1) - zb + [(\alpha kz^2(k+1) - zb)^2 + 4kz(q + kz^2\alpha^2k - b\alpha z)]^{1/2}}{2kz^2},$$

$$z = \frac{b \pm (b^2 - 4k)^{1/2}}{2\alpha k}$$ (see Remark 3.2.2.1′, p. 270);

(c) if $k + q > 2$, $q < 1$, and $\nu = -1$, then $\alpha = (1-q)^{-1}$, $\beta = (k-1)(1-q)^{-1} - 1$, $l_0 = b/2 - [b^2/4 - (dR/d\Theta|_{\Theta=1})]^{1/2}$, and $b \geqslant 2(dR/d\Theta|_{\Theta=1})^{1/2}$, where $R(\Theta) = \nu k\Theta^{k+q-1}(1-\Theta^\mu)$.

Function $g(\tau, t, 0)$ has the form

$$g(\tau, t, 0) = -\int_0^\tau \exp(I(\xi))\left[\int_a^\xi \tilde{f}(\eta, t)\exp(-I(\eta))\,d\eta\right]d\xi, \tag{3.8.0.40}$$

with

$$I(\eta) = \int^{\eta} \left(\frac{b}{k} \chi^{1-k} - 2 \frac{d}{d\xi} \ln \frac{d\chi^k}{d\xi} \right) d\xi$$

and function $\tilde{f}$ defined thus:

$$\tilde{f}(\xi, t) = \frac{1}{A^{k/\mu}(-\varphi, t)\,\beta^2 \lambda(-\varphi, t)\,(d\chi^k/d\xi)}$$

$$\times \Big\{ -\xi \frac{d\chi}{d\xi} A^{1/\mu}(-\varphi, t) \frac{1}{\beta} \left(\frac{d\beta}{dt} + 2\beta_1 \frac{d\varphi}{dt} \right)$$

$$+ \xi \frac{d^2\chi^k}{d\xi^2} A^{k/\mu}(-\varphi, t) \left[\beta \frac{\partial \lambda}{\partial x}\Big|_{x=-\varphi} + 4\beta_1 \lambda(-\varphi, t) \right]$$

$$+ A^{k/\mu}(-\varphi, t)\, 2\beta_1 \lambda(-\varphi, t) \frac{d\chi^k}{d\xi}$$

$$+ \beta \frac{\partial \lambda}{\partial x}\Big|_{x=-\varphi} A^{k/\mu}(-\varphi, t) \frac{d\chi^k}{d\xi}$$

$$+ 2\lambda(-\varphi, t) \frac{\partial A^{k/\mu}}{\partial x}\Big|_{x=-\varphi} \beta \frac{d\chi^k}{d\xi}$$

$$+ \nu \frac{\partial \gamma^2}{\partial x}\Big|_{x=-\varphi} \frac{\xi}{\beta} \chi^q (1 - \chi^\mu) A^{q/\mu+1}(-\varphi, t)$$

$$+ \delta(-\varphi, t) A^{1/\mu}(-\varphi, t)\, \xi \frac{d\chi}{d\xi} \frac{2\beta_1}{\beta}$$

$$+ \frac{\partial \delta}{\partial x}\Big|_{x=-\varphi} \xi A^{1/\mu}(-\varphi, t) \frac{d\chi}{d\xi} \Big\}.$$

If $k + q > 2$ and $q \geqslant 1$, then the lower limit q in the inner integral in (3.8.0.40) is zero; if $q < 1$ and $k + q = 2$, then $a = 0$.

If $q < 1$ and $k + q > 2$, then $a = \infty$.

For function $g(\tau, t, 0)$ the following estimates hold true:

$$g = O(\tau^2) \text{ as } \tau \to 0;$$
$$g = O(\tau^2 \exp(-l_0 \tau k)) \text{ as } \tau \to \infty .$$

Function $\beta_1(t, \varepsilon)$ is defined thus: $\beta_1(t, \varepsilon) = \tilde{\beta}_1$ if $a = 0$, $\beta_1 = \lim_{\tau \to 0} \tilde{\beta}_1$ if $a = +\infty$,

$$\tilde{\beta}_1 \overset{\text{def}}{=} \Big[\frac{2}{\beta} A^{1/\mu}(-\varphi, t) \left(\frac{d\varphi}{dt} - \delta(-\varphi, t) \right) I_2$$
$$- 4\lambda(-\varphi, t) I_4 A^{k/\mu}(-\varphi, t)$$
$$- 2\lambda(-\varphi, t) I_3 A^{k/\mu}(-\varphi, t) \Big]^{-1}$$

$$\times \Big[A^{1/k}(-\varphi,\ t) \left(\frac{\partial \delta}{\partial x}\Big|_{x=-\varphi} - \frac{1}{\beta}\frac{d\beta}{dt} \right) I_2$$

$$+ A^{k/\mu}(-\varphi,\ t)\beta \frac{\partial \lambda}{\partial x}\Big|_{x=-\varphi} I_4$$

$$+ \beta \left(\frac{\partial \lambda}{\partial x}\Big|_{x=-\varphi} A^{k/\mu}(-\varphi,\ t) \right.$$

$$\left. + 2\lambda(-\varphi,\ t) \frac{\partial A^{k/\mu}}{\partial x}\Big|_{x=-\varphi} \right) I_3$$

$$+ \frac{\partial \gamma^2}{\partial x}\Big|_{x=-\varphi} \frac{1}{\beta} A^{q/k+1}(-\varphi,\ t)\, I_5 \Big],$$

where if $a = 0$, the integrals are evaluated according to the formulas

$$\begin{aligned}
I_2 &= \int_0^\tau \frac{d\chi^k}{d\xi} \exp\left(-\frac{b}{k}\int^\xi \chi^{1-k}\, d\xi' \right) \xi \frac{d\chi}{d\xi}\, d\xi, \\
I_3 &= \int_0^\tau \left(\frac{d\chi^k}{d\xi} \right)^2 \exp\left(-\frac{b}{k}\int^\xi \chi^{1-k}\, d\xi' \right) d\xi, \\
I_4 &= \int_0^\tau \frac{d\chi^k}{d\xi} \exp\left(-\frac{b}{k}\int^\xi \chi^{1-k}\, d\xi' \right) \xi \frac{d^2\chi^k}{d\xi^2}\, d\xi, \\
I_5 &= \int_0^\tau \frac{d\chi^k}{d\xi} \exp\left(-\frac{b}{k}\int^\xi \chi^{1-k}\, d\xi' \right) \xi\chi^q (1-\chi^\mu)\, d\xi,
\end{aligned} \tag{3.8.0.41}$$

while if $a = +\infty$, the integrals are evaluated according to the formulas

$$\begin{aligned}
I_2 &= \int_0^\tau \exp\left(\frac{b}{k}\int^\xi \chi^{1-k}\, d\xi' \right) \left(\frac{d\chi^k}{d\xi} \right)^{-2} \Big[\int_\xi^\infty \frac{d\chi^k}{d\xi'} \\
&\quad \times \exp\left(-\frac{b}{k}\int^{\xi'} \chi^{1-k}\, d\xi \right) \xi' \frac{d\chi}{d\xi'}\, d\xi' \Big] d\xi, \\
I_3 &= \int_0^\tau \exp\left(\frac{b}{k}\int^\xi \chi^{1-k}\, d\xi' \right) \left(\frac{d\chi^k}{d\xi} \right)^{-2} \Big[\int_\xi^\infty \left(\frac{d\chi^k}{d\xi} \right)^2 \\
&\quad \times \exp\left(-\frac{b}{k}\int^{\xi'} \chi^{1-k}\, d\xi \right) d\xi' \Big] d\xi, \\
I_4 &= \int_0^\tau \left(\frac{d\chi^k}{d\xi} \right)^{-2} \exp\left(\frac{b}{k}\int^\xi \chi^{1-k}\, d\xi' \right) \Big[\int_\xi^\infty \frac{d\chi^k}{d\xi'}
\end{aligned} \tag{3.8.0.42}$$

$$\times \exp\left(-\frac{b}{k}\int^{\xi'}\chi^{1-k}\,d\xi\right)\xi'\frac{d^2\chi^k}{d\xi'^2}\,d\xi'\Big]\,d\xi,$$

$$I_5=\int_0^{\tau}\left(\frac{d\chi^k}{d\xi}\right)^{-2}\exp\left(\frac{b}{k}\int^{\xi}\chi^{1-k}\,d\xi'\right)\Big[\int_{\xi}^{\infty}\frac{d\chi^k}{d\xi}$$

$$\times \exp\left(-\frac{b}{k}\int^{\xi'}\chi^{1-k}\,d\xi\right)\xi'\chi^q(1-\chi^\mu)\,d\xi'\Big]\,d\xi.$$

Let us describe the case where function $A\ (x,\ t)$ does not satisfy Eq. (3.8.0.37) and $k>1$, $q>0$, $k+q\geqslant 2$, and $\mu>0$. Then the asymptotic to within $O\ (\varepsilon^2)$ solution to problem (3.8.0.36) exists and has the form

$$u(x,\ t,\ \varepsilon)=A^{1/\mu}(x,\ t)\,\chi\,(\xi+O\,(\varepsilon^2)).$$

Function χ solves problem (3.8.0.38), and functions β and φ are defined via system (3.8.0.39). In view of the fact that function $g\ (\xi,\ t,\ x,\ \varepsilon)$ can not be expanded in a Taylor series at point $\xi=\tau$ in such a manner that the uniform (in τ) estimate of the remainder term remains valid, $g\ (\xi,\ t,\ x,\ 0)$ has the form

$$g(\xi,\ t,\ x,\ 0)=-\int_0^{\xi}\exp\,(I\,(\xi))\Big[\int_a^{\xi}\widetilde{f}\,(\eta)\exp\,(-I\,(\eta))\,d\eta\Big]\,d\xi,$$

$$I\,(\eta)=\int_0^{\eta}\left(\frac{b}{k}\chi^{1-k}-2\frac{\partial}{\partial\xi}\ln\frac{\partial\chi^k}{\partial\xi}\right)d\xi;$$

as $\xi\to\infty$, the following estimate holds true uniformly in x, t, and ε:

$$g(\xi,\ t,\ x,\ 0)\sim\frac{k}{l_0}\left(\exp\frac{l_0\xi}{k}\right)\omega_0(x,\ t),$$

where

$$\omega_0(x,\ t)=\lim_{\tau\to\infty}\frac{d\chi}{d\xi}\,g\,(\xi,\ t,\ x,\ 0);$$

finally,

$$g=O\,(\xi^2)\ \text{ as }\ \xi\to 0.$$

Function $\widetilde{f}$ is determined by the expression

$$\widetilde{f}\,(\xi,\ t)=\frac{1}{A^{k/\mu}(-\varphi,\ t)\,\beta^{2\lambda}(-\varphi,\ t)\,(d\chi^k/d\xi)}$$

$$\times\left\{-\xi\frac{d\chi}{d\xi}A^{1/\mu}(-\varphi,\ t)\frac{1}{\beta}\left(\frac{d\beta}{dt}+2\beta_1\frac{d\varphi}{dt}\right)\right.$$

$$+\xi \frac{d^2\chi^k}{d\xi^2} A^{k/\mu}(-\varphi, t)\left[\beta \frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi} + 4\beta_1\lambda(-\varphi, t)\right]$$

$$+ A^{k/\mu}(-\varphi, t)\, 2\beta_1\lambda(-\varphi, t)\frac{d\chi^k}{d\xi}$$

$$+\beta Ak/\mu(-\varphi, t)\frac{d\chi^k}{d\xi}\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}$$

$$+2\lambda(-\varphi, t)\,\beta \frac{d\chi^k}{d\xi}\frac{\partial A^{k/\mu}}{\partial x}\Big|_{x=-\varphi}$$

$$+\nu \frac{\partial\gamma^2}{\partial x}\Big|_{x=-\varphi}\frac{\xi}{\beta}\chi^q(1-\chi^\mu)\,A^{q/\mu+1}(-\varphi, t)$$

$$+\delta(-\varphi, t)\,A^{1/\mu}(-\varphi, t)\,\xi\frac{d\chi}{d\xi}\frac{2\beta_1}{\beta}$$

$$+\xi A^{1/\mu}(-\varphi, t)\frac{d\chi}{d\xi}\frac{\partial\delta}{\partial x}\Big|_{x=-\varphi}$$

$$+\chi(\xi)\left[-\frac{\partial A^{1/\mu}(-\varphi, t)}{\partial t} + \delta(-\varphi, t)\,\frac{\partial A^{1/\mu}(-\varphi, t)}{\partial x}\right]\Bigg\}.$$

If $q \geqslant 1$, $k + q > 2$, then the lower limit a in the inner integral is zero.

If $q < 1$ and $k + q = 2$, then $a = 0$.

If $q < 1$ and $k + q > 2$, then $a = +\infty$.

Function $\beta_1(t, \varepsilon)$ is defined thus: $\beta_1(t, \varepsilon) = \tilde{\beta}_1$ if $a = 0$, $\beta_1(t, \varepsilon) = \lim\limits_{\tau\to 0}\tilde{\beta}$ if $a = +\infty$,

$$\tilde{\beta}_1 \overset{\text{def}}{=} \Big\{2A^{1/\mu}(-\varphi, t)\,I_2\left[\frac{d\varphi}{dt} - \delta(-\varphi, t)\right]\beta^{-1}$$

$$-4\lambda(-\varphi, t)\,A^{k/\mu}(-\varphi, t)\,I_4 - 2\lambda(-\varphi, t)\,I_3A^{k/\mu}(-\varphi, t)\Big\}^{-1}$$

$$\times\Big\{A^{1/\mu}(-\varphi, t)\,I_2\left(\frac{\partial\delta}{\partial x}\Big|_{x=-\varphi} - \frac{1}{\beta}\frac{d\beta}{dt}\right)$$

$$+A^{k/\mu}(-\varphi, t)\,I_4\beta\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}$$

$$+\beta I_3\Big[A^{k/\mu}(-\varphi, t)\frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi}$$

$$+2\lambda(-\varphi, t)\frac{\partial A^{k/\mu}(-\varphi, t)}{\partial x}\Big|_{x=-\varphi}\Big]$$

$$+\frac{\nu}{\beta}I_5A^{q/\mu+1}(-\varphi, t)\frac{\partial\gamma^2}{\partial x}\Big|_{x=-\varphi}\Big\}.$$

If $k + q > 2$ and $q \geqslant 1$ or $k + q = 2$ and $q < 1$, then $a = 0$ and the integrals I_m are evaluated by formulas (3.8.0.41).

If $q < 1$ and $k + q > 2$, then $a = +\infty$ and the integrals I_m are evaluated by formulas (3.8.0.42).

1.9. Quasilinear hyperbolic equations:

$$\varepsilon \frac{\partial u}{\partial t} + \varepsilon^2 \frac{\partial^2 u}{\partial t^2} - \varepsilon^2 \frac{\partial}{\partial x} \left(\lambda (x, t) K (u) \frac{\partial u}{\partial x}\right) - \mathcal{F} (u) = 0, \qquad (3.8.0.43)$$

$\mathcal{F} (a_i) = 0,\ i = 0,\ 1,\ 0 < a_0 < a_1,\ K (u) > 0,\ dK/du \neq 0$

for $u \in (a_0, a_1)$.

The boundary conditions:

$$u|_{x\to-\infty} = a_0, \quad u_{x\to\infty} = a_1.$$

The basic formulas:

$$u (x, t, \varepsilon) = W (S/\varepsilon + \varepsilon g (S/\varepsilon, t, x) + O (\varepsilon^2)), \qquad (3.8.0.44)$$

$$S (x, t, \varepsilon) = \beta (t) (x + \varphi (t)) + \beta_1 (t) (x + \varphi (t))^2, \qquad (3.8.0.45)$$

$$b \frac{dW}{d\xi} - \frac{d}{d\xi} \left((K (W) - b^2) \frac{dW}{d\xi}\right) - \mathcal{F} (W) = 0,\ \xi = \tau + \varepsilon g, \qquad (3.8.0.46)$$

$$a_0 < W < a_1,\ W|_{\xi=0} = a_0,\ W_{\xi\to+\infty} = a_1,\ (K (W) - b^2) \frac{dW}{d\xi}\Big|_{\xi=0} = 0,$$

$$\frac{d\varphi}{dt} = b \sqrt{\lambda (-\varphi, t)},\ \beta (t) = \sqrt{\lambda^{-1} (-\varphi, t)}, \qquad (3.8.0.47)$$

$$b_0 \frac{d\Theta}{d\zeta} - \frac{d^2\Theta}{d\zeta^2} - R (\Theta) = 0, \text{ where } R (\Theta) = (K (\Theta) - b^2) \mathcal{F} (\Theta), \qquad (3.8.0.48)$$

$$g (\tau, t, 0) = - \int_0^\tau (\exp I (\xi)) \left[\int_a^\xi f (\eta, t) \exp [- I (\eta)] \, d\eta \right] d\xi, \qquad (3.8.0.49)$$

$$I (\xi) = -2 \ln \left[K (W) \frac{dW}{d\xi} - b^2 \frac{dW}{d\xi} \right] + \int^\xi b (K (W) - b^2)^{-1} \, d\xi,$$

$$\begin{aligned} f = {} & \left[\frac{dW}{d\xi} (K (W) - b^2) \right]^{-1} \left\{ \frac{dW}{d\xi} \frac{\xi}{\beta} \left(\frac{d\beta}{dt} + 2\beta_1 \frac{d\varphi}{dt} \right) \right. \\ & + 2 \frac{d^2 W}{d\xi^2} \xi \frac{d\varphi}{dt} \left(2\beta_1 \frac{d\varphi}{dt} + \frac{d\beta}{dt} \right) + \frac{dW}{d\xi} \left[\frac{d}{dt} \left(\beta \frac{d\varphi}{dt} \right) \right. \\ & \left. + \frac{d\beta}{dt} \frac{d\varphi}{dt} + 2\beta_1 \left(\frac{d\varphi}{dt} \right)^2 \right] - \xi \frac{d}{d\xi} \left(K (W) \frac{dW}{d\xi} \right) \left(4\beta_1 \lambda (- \varphi, t) \right. \\ & \left. \left. + \beta \frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi} \right) - K (W) \frac{dW}{d\xi} \left(2\beta_1 \lambda (- \varphi, t) + \beta \frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi} \right) \right\}. \end{aligned}$$

$$g = O (\tau^2) \text{ as } \tau \to 0,$$

$$g = O (\tau^2 \exp [- l_0 \tau (K (a_1) - b^2)^{-1}]) \text{ as } \tau \to \infty. \qquad (3.8.0.50)$$

$$\tilde{\beta}_1 \stackrel{\text{def}}{=} \left\{ \frac{2}{\beta} \frac{d\varphi}{dt} I_0 + 4I_4 \left(\frac{d\varphi}{dt}\right)^2 - 4\lambda(-\varphi,\ t)\, I_3 + 2I_1 \left(\frac{d\varphi}{dt}\right)^2 - 2\lambda(-\varphi,\ t)\, I_2 \right\}^{-1} \left\{ -\frac{1}{\beta}\frac{d\beta}{dt} I_0 - 2\frac{d\varphi}{dt}\frac{d\beta}{dt} I_4 - \left[\frac{d}{dt}\left(\beta \frac{d\varphi}{dt}\right) + \frac{d\beta}{dt}\frac{d\varphi}{dt}\right] I_1 + I_3 \beta \frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi} + \beta I_2 \frac{\partial\lambda}{\partial x}\Big|_{x=-\varphi} \right\}. \tag{3.8.0.51}$$

Here

$$I_0 = \int_0^\infty \xi \left(\frac{dW}{d\xi}\right)^2 M_1(\xi)\, d\xi,$$

$$I_1 = \int_0^\infty \left(\frac{dW}{d\xi}\right)^2 M_1(\xi)\, d\xi,$$

$$I_2 = \int_0^\infty \left(\frac{dW}{d\xi}\right)^2 K(W)\, M_1(\xi)\, d\xi, \tag{3.8.0.52}$$

$$I_3 = \int_0^\infty \frac{dW}{d\xi}\, \xi \frac{d}{d\xi}\left(K(W) \frac{dW}{d\xi}\right) M_1(\xi)\, d\xi,$$

$$I_4 = \int_0^\infty \xi \frac{dW}{d\xi} \frac{d^2W}{d\xi^2} M_1(\xi)\, d\xi,$$

where

$$M_1(\xi) = (K(W) - b^2)\ \exp\left[-\int^{\xi} b\,(K(W) - b^2)^{-1}\, d\xi\right],$$

or[16]

$$I_0 = \int_0^\tau M(\xi) \left(\int_\xi^\infty \left(\frac{dW}{d\xi'}\right)^2 \xi' M_1(\xi')\, d\xi'\right) d\xi,$$

$$I_1 = \int_0^\tau M(\xi) \left(\int_\xi^\infty \left(\frac{dW}{d\xi'}\right)^2 M_1(\xi')\, d\xi'\right) d\xi,$$

$$I_2 = \int_0^\tau M(\xi) \left(\int_\xi^\infty \left(\frac{dW}{d\xi'}\right)^2 M_1(\xi') K(W)\, d\xi'\right) d\xi, \tag{3.8.0.53}$$

[16] Below we will discuss which formulas for I, (3.8.0.52) or (3.8.0.53), should be used in each specific case.

$$I_3 = \int_0^\tau M(\xi)\left(\int_\xi^\infty \frac{dW}{d\xi'}\frac{d}{d\xi'}\left(K(W)\frac{dW}{d\xi'}\right)M_1(\xi')\,d\xi'\right)d\xi,$$

$$I_4 = \int_0^\tau M(\xi)\left(\int_\xi^\infty \xi'\frac{d^2W}{d\xi'^2}M_1(\xi')\,d\xi'\right)d\xi,$$

where

$$M(\xi)=(K(W)-b^2)^{-2}\exp\left(-\int^\xi b(K(W)-b^2)^{-1}\,d\xi\right).$$

A.1. Suppose that $1<k<2$, $q>0$, $\mu>0$, and $K(u)\sim \tilde{K}(a_0)+\mu(u-a_0)^{k-1}$ and $\mathcal{F}(u)\sim(u-a_0)^q$ as $u\to a_0$, with $d\mathcal{F}/du\,|_{u=a_0}<0$. The asymptotic solution to problem (3.8.0.43) has the form (3.8.0.44). Function $W(\xi)$ is a partial solution to problem (3.8.0.46), function S has the form (3.8.0.45), and functions β and φ can be found by solving system (3.8.0.47).

The asymptotic expansion of W as $\xi\to 0$ has the form

$$W=a_0+C_1\xi^\alpha+o(\xi^\alpha),\quad \alpha=(k-1)^{-1}, \tag{3.8.0.54}$$

and the following estimate holds true:

$$W\sim a_1-\exp(-|l_1(K(a_1)-b^2)^{-1}|\xi)\ \text{as}\ \xi\to\infty. \tag{3.8.0.55}$$

If $q\geqslant 1$ and $k+q=2$ or $q<1$ and $k+q=2$, then $l_1=-b/2+[b^2/4-(dR/d\Theta\,|_{\Theta=a_1})]^{1/2}$.

If $q<1$ and $k+q>2$, then $l_1=b/2-[b^2/4-(dR/d\Theta\,|_{\Theta=a_1})]^{1/2}$.

If $k+q=2$, then $b>2\,(dR/d\Theta\,|_{\Theta=a_0})^{1/2}$ (see Remark 3.2.2.1′).

If $k+q>2$, then $b=b_0$ is the Zeldovich constant.

The necessary condition for the existence of a solution of the (3.8.0.44) type is that $b^2=K(a_0)$. Function $g(\tau,t,\varepsilon)$ is defined via (3.8.0.49) and possesses the estimate (3.8.0.50).

If $q>1$, $k+q>2$, or $q<1$, $k+q=2$, then the lower limit a in the inner integral in (3.8.0.49) is zero.

If $q<1$ and $k+q>2$, then $a=+\infty$.

Function $\beta_1\equiv\tilde{\beta}_1$ is defined via (3.8.0.51), where

(a) if $q\geqslant 1$, $k+q>2$, or $q<1$, $k+q=2$, then the integrals are evaluated via formulas (3.8.0.52); and

(b) if $q<1$, $k+q\geqslant 2$, $\beta_1=\lim\limits_{\tau\to 0}\tilde{\beta}_1$ then the integrals are evaluated via formulas (3.8.0.53).

A.2. Suppose that $\mu < 0$, $q > 0$, $k + q \geqslant 2$, $K(u) = \tilde{K}(a_0) + \mu (u - a_0)^{k-1}$, $\mathcal{F}(u) \sim (u - a_0)^q$, and $d\mathcal{F}/du\,|_{u=a_1} < 0$. The asymptotic solution to problem (3.8.0.43) has the form (3.8.0.44). Function $W(\xi)$ solves problem (3.8.0.46), function S has the form (3.8.0.45), and functions β and φ can be found by solving system (3.8.0.47).

The asymptotic expansion of W as $\xi \to 0$ has the form

$$W = a_0 + C_1 \xi^\alpha + o(\xi^\alpha), \quad \alpha = (1 - q)^{-1},$$

and the following estimate holds true:

$$W \sim a_1 - \exp(-|\, l_1 (K(a_1) - b^2)^{-1}\, |\, \xi) \quad \text{as} \quad \xi \to \infty.$$

The necessary condition for the existence of a solution of the (3.8.0.44) type is that $b^2 = K(a_0)$. Functions g and β_1 are defined in the same way as in Subsection *A.1*, that is, by (3.8.0.49) when $a = \infty$ and by (3.8.0.51) when $a = 0$.

A.3. Suppose that $q > 0$, $\mu < 0$, $K(u) = \tilde{K}(a_1) - \mu (a_1 - u)^{k-1}$, $\mathcal{F}(u) \sim (a_1 - u)^q$, and $d\mathcal{F}/du\,|_{u=a_0} > 0$. The asymptotic solution to problem (3.8.0.43) has the form (3.8.0.44). Function $W(\xi)$ solves problem (3.8.0.46), where $a_0 \leqslant W \leqslant a_1$, $W\,|_{\xi \to -\infty} \to a_0$, $W\,|_{\xi \to 0} = a_1$, and $(K(W) - b^2)(dW/d\xi\,|_{\xi=0}) = 0$.

If $k + q = 2$, then $b > 2\,(|\,dR/dW\,|_{W=a_1}|)^{1/2}$.

If $k + q > 2$, then $b = b_0$ is the Zeldovich constant in Eq. (3.8.0.48).

The asymptotic expansion of W has the form

$$W = a_1 - C_1 \xi^\alpha + o(\xi^\alpha), \quad \alpha = (k - 1)^{-1}, \qquad (3.8.0.56)$$

and the following estimate holds true:

$$W \sim a_0 + \exp(|\,l_1 (K(a_0) - b^2)^{-1}|\,\xi) \text{ as } \xi \to -\infty; \quad (3.8.0.57)$$

here if $q \geqslant 1$, $k + q > 2$ or $q < 1$, $k + q = 2$, then $l_1 = -b/2 + [b^2/4 - (dR/d\Theta\,|_{\Theta=a_0})]^{1/2}$.

The necessary condition for the existence of a solution of the (3.8.0.44) type is that $b^2 = K(a_1)$. Function $g(\tau, t, \varepsilon)$ is defined via (3.8.0.49) when $a = -\infty$, while in formulas (3.8.0.52) and (3.8.0.53) the integration limit ∞ should be replaced with $-\infty$.

A.4. Suppose that $\mu > 0$, $0 < q < 1$, $k + q > 2$, $K(u) = \tilde{K}(a_1) - \mu (a_1 - u)^{k-1}$, $\mathcal{F}(u) \sim (a_1 - u)^q$, and $d\mathcal{F}/du\,|_{u=a_0} > 0$.

The asymptotic solution to problem (3.8.0.43) has the form (3.8.0.44). Function $W(\xi)$ solves problem (3.8.0.46), where $a_0 \leqslant W < a_1$, $W\,|_{\xi \to -\infty} \to a_0$, $W\,|_{\xi=0} = a_1$, $(K(W) - b^2)(dW/d\xi\,|_{\xi=0}) = 0$, and $b > 2\,(|\,dR/dW\,|_{W=a_0}|)^{1/2}$.

The asymptotic expansion of W as $\xi \to 0$ has the form

$$W \simeq a_1 - C_1 \xi^\alpha + o(\xi^\alpha), \quad \alpha = (1 - q)^{-1},$$

and estimate (3.8.0.54) holds true. Functions g ad β_1 are defined in the same way as in Subsection *A.3*. The necessary condition for the existence of a solution of the (3.8.0.44) type is that $b^2 = K(a_1)$.

References

3.1. A.R. Neureuter, *Proc. IEEE* **71**: 149-162 (1983).
3.2. L.I. Sedov, *Similarity and Dimensional Methods in Mechanics* (Moscow: Mir Publishers, 1982).
3.3. V.P. Maslov, V.G. Danilov, and K.A. Volosov, *Mathematical Modeling of Heat and Mass Transfer Processes: The Evolution of Dissipative Structures* (With an Addition by N.A. Kolobov) (Moscow: Nauka, 1987) (in Russian).
3.4. N.A. Kolobov and M.M. Krymko. in: *Review of Electronic Technology*, Ser. 2 (PP) (Moscow: Central Scientific Research Institute "Elektronika", 1978) (in Russian).
3.5. N.A. Kolobov and M.M. Samokhvalov, *Diffusion and Oxidation in Semiconductors* (Moscow: Vysshaya shkola, 1980) (in Russian).
3.6. N.A. Kolobov, *Fundamentals of the Technology of Electronic Devices* (Moscow: Metallurgiya, 1975) (in Russian).
3.7. N.S. Enikolopov and S.A. Vol'fson, *Calculations of Highly Effective Processes* (Moscow: Khimiya, 1980) (in Russian).
3.8. J.W. Rouse, C.R. Helms, B.E. Deal, and R.R. Razouk, *J. Electrochem. Soc.* **131**, No. 4: 887-894 (1984).
3.9. L.A. Zhukova, *The Theory of Static and Dynamic Precipitation and Coprecipitation of Ions* (Moscow: Energoizdat, 1981) (in Russian).
3.10. G. Krtschman, *Schweisstechnik* (DDR) **19**, No. 5: 198-202 (1969).
3.11. V.I. Polezhaev and N.A. Verezub, in: *Gagarin Readings on Astronautics and Aviation in 1983 and 1984* (Moscow: Nauka, 1985): pp. 300-302 (in Russian).
3.12. N.A. Avdonin, *Mathematical Description of Crystallization Processes* (Riga: Zinatne, 1980) (in Russian).
3.13. V.P. Il'in, *Numerical Methods for Solving Electrophysics Problems* (Moscow: Nauka, 1985) (in Russian).
3.14. V.V. Andrianov, M.B. Perizh, and S.I. Kopylov, *Preprint IVT AN SSSR No. 4-070* (Moscow, 1980) (in Russian).
3.15. R.G. Mints and A.L. Rakhmanov, *Instabilities in Superconductors* (Moscow: Nauka, 1984) (in Russian).
3.16. M. Tunkham, *Introduction to Superconductivity* (New York: McGraw-Hill, 1979).
3.17. V.V. Shmidt, *Introduction to Semiconductor Physics* (Moscow: Nauka, 1982) (in Russian).
3.18. V.A. Fock, *Trudy Gos. Opt. Inst. (Leningrad)* **4**, No. 34 (1926).
3.19. *New Scientist* 93, 1288 (1982).
3.20. *Electric Design* 32, No. 10: 259-274 (1984).
3.21. I.S. Aranson and M.I. Rabinovich, *Preprint IPF AN SSSR No. 152* (Gorki, 1987) (in Russian).
3.22. A.V. Gaponov and M.I. Rabinovich, in: *Proc. Intern. Conf. on Plasma Physics* (Kiev, 1987).
3.23. A.B. Ezerskii, M.I. Rabinovich, V.P. Reutov, and I.M. Starobinets, *Soviet Phys. JETP* 64, No. 6: 1228-1236 (1986).
3.24. V.E. Zakharov, V.S. L'vov and S. S. Starobinets, *Soviet Physics Uspekhi*: 17, No. 6, 896 (1975).
3.25. J.D. Murray, *Lectures on Nonlinear-Differential-Equation Models in Biology* (Oxford: Clarendon Press, 1977).

3.26. A.N. Kolmogorov, I.G. Petrovskii, and N.S. Piskunov, *Bull. MGU*, Sec. A I, No. 6: 1-25 (1937).
3.27. P.G. Fife and J.B. McLeog, *Arch. Rational Mech. Anal.* **65**, No. 3: 336-361 (1983).
3.28. D.G. Aronson and H.F. Weinberger, in: *Lecture Notes in Mathematics* (Colostein ed.), vol. 449 (New York: Springer, 1975): pp. 5-49.
3.29. R.A. Fisher, *Ann. Eugenics* **7**, No. 17: 335-369 (1937).
3.30. V.S. Berman, *Dokl. Akad. Nauk SSSR* **242**, No. 2: 265-267 (1979).
3.31. V.G. Danilov, N.A. Kolobov, V.P. Maslov, and K.A. Volosov, *Sov. Math. Dokl.* **33**, No. 2: 517-521 (1986).
3.32. G. Sansone, *Equazioni differenziali nel campo reale*, Part 2 (Bologna, 1949).
3.33. S. Lefschetz, *Differential Equations: Geometric Theory* (New York: Interscience, 1957).
3.34. E.V. Tolubinskii, *The Theory of Transfer Processes* (Kiev: Naukova dumka, 1969) (in Russian).
3.35. E.I. Andriankin, *Soviet Phys. JETP* **35**, No. 2: 295-299 (1959).
3.36. E.I. Andriankin, *Zh. Tekh. Fiz.* **29**, No. 11: 1368-1372 (1959).
3.37. V.I. Krinsky (ed.), *Self-Organization: Autowaves and Structures Far from Equilibrium* (Berlin: Springer, 1984).
3.38. H. Frohlich, *MIT Neurosci. Res. Progr. Bull.* **67**, No. 15: 117-121 (1977).
3.39. R.K. Bullough and Ph. Candrey (eds.), *Solitons* (Berlin: Springer, 1980).
3.40. L.K. Martinson, in: *Mathematical Modeling: Processes in Nonlinear Media* (Moscow: Nauka, 1986) pp. 279-309 (in Russian).
3.41. K.A. Volosov, *Inzh.-Fiz. Zh.* **26**, No. 5: 929-930 (1981).
3.42. K.A. Volosov and I.A. Fedotov, *USSR Comput. Maths. Math. Phys.* **23**, No. 5, 147-150 (1983).
3.43. Kohei Uchiyama, *J. Math. Kyoto Univ.* **3**, No 18 : 453-508 (1978).

NAME INDEX

SUBJECT INDEX

TO THE READER

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:
Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820

Also from Mir Publishers

THE THEORY OF CHOICE AND DECISION-MAKING

I. MAKAROV, Mem. USSR Acad. Sc., *et al.*

The volume covers a broad spectrum of mathematical theories pertinent to the present-day understanding of choice and decision-making. It offers techniques and theory, and is full of numerous applications of decision-making methodologies and thus is intended to serve both as a guide for a wide range of analysts and experts closely involved in the mathematical modelling of decision processes and as a textbook for students.

Contents. Introduction. Theoretical Background for the Choice of Alternatives. Binary Relations. Choice Functions. Binary Relations on E_m. Coordinate Relations. Decomposition of Choice Functions. Procedures and Algorithms for Decision-Making. Expert Procedures for a Decision-Making. Methods of Processing Expert Information. Forming the Initial Set of Alternatives. Choice Problems. Probabilistic Characteristics of the Cardinal Number of Ω^R. Utility Functions in Choice Problems. Choice Problems with a Given Optimality Principle. Optimal Control with Multiple Criteria. Control with Multiple Criteria. Discrete Multiple Criteria Problems. Continuous Time Multiple-Criteria Problems. Markov Models of Decision-Making. Applied Multiple-Criteria Problems of Optimal Control. Bibliography. Subject Index.